

Dummy Variable Regression Models and Analysis of Responses in Diagnostic Screening Tests with Replications

I. C. A. Oyeka, E. U. Umeh*

Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria

Abstract This paper develops and presents use of dummy variable regression models in the analysis of results from diagnostic screening tests or clinical trials with several case-control subject response options or categories, with replications in each case-control response category or combination. The proposed method can also be used in comparing differences between proportions of case and control subjects responding positive under various case-control response categories or options. Test statistics are developed for testing the null hypotheses that the regression model is an adequate representation of the differences in scores or responses between pairs of case and control subjects in diagnostic screening tests and for comparing the responses of cases and controls in the paired case-control subject response categories. Test statistics are also developed for use in pairwise comparisons of the effect of scores by case-control subject at various response categories. Corresponding probabilities, odds and odds-ratios of positive responses are also estimated using the resulting regression coefficients. The proposed method is illustrated with some data.

Keywords Dummy Variables, Standard drug, New drug, Probabilities, Odds and Odds Ratio, Replications

1. Introduction

Often in controlled comparative either prospective, retrospective or cross – sectional study involving matched samples of subjects or patients, the response of a subject or patient to a predisposing factor in a retrospective study or to a condition or treatment in a prospective study may be dichotomous but more often much finer than simply dichotomous. These responses may be of such form as yes or no, present or absent, alive or dead, positive or negative, etc. But there may also be several other possible response options. For example, in a retrospective study where the predisposing factor of a certain conditions is the respondents drinking habit, a subject may be classified as a teetotler, a light drinker or a heavy drinker etc. in a prospective study involving same conditions or test, a patient may be classified as recovered, much improved, improved, no change, worse or dead. A treatment or drug may be graded as very effective, effective or ineffective etc.

Furthermore there may not be only one pair but several pairs of case control subject combinations and other such situations. Thus specifically suppose we have a total random sample of n - matched pairs of patients or subjects

matched on certain characteristics to be exposed to two treatments, drugs, tests; suppose n_{ij} pairs of these subject pairs are in the i^{th} response category by case and j^{th} response category by the control subject, for $i, j = 1, 2, \dots, c$ responses. In other words, suppose, one member of each pair of subjects or patients in the $(i, j)^{th}$ case control response combination is exposed to one under control of the two treatments, drugs, tests or procedures of interest and the other member of pair is exposed to the remaining (case, new drug) treatment, drug, test or procedure T_2 . Furthermore, suppose that the response of members of each pair of case-control subjects or patients are more than dichotomous but numbering some c ($c \geq 2$) possible mutually exclusive response options or categories.

In Table 1, there is no replication, that is, there is only one pair of case-control subjects. Of the $n = n_{..}$ pairs of case-control subjects used in the study, $n_{i.}$ and $n_{.j}$ are subjects respectively in the i^{th} response category for case and j^{th} response category for control while n_{ij} pairs are in the $(i, j)^{th}$ pair of case- control subject response category for $i, j = 1, 2, \dots, c$ response categories.

Results from these types of study may be analyzed using several methods for example Backpar (1966); Grizzle, Starmer and Kosh (1959), Ireland, Ku and Kullback (1969); Stuart (1955) and Maxwell (1970), if $c \leq 3$ or the more generalised method can be used. We present an alternative method of analysis in terms of probabilities, odds and odds-ratios of occurrence of outcomes using dummy variable regression techniques.

* Corresponding author:

editus2002@yahoo.com (E. U. Umeh)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

Table 1. Format for presentation of results in diagnostic screening tests with “c” outcomes

State of Nature (Control, Standard, T_1)								
Case (New drug, T_2)	1	2	.	.	.	c	.	.Total (n_{1i})
1	n_{11}	n_{12}	.	.	.	n_{1c}	.	$n_{1.}$
2	n_{21}	n_{22}	.	.	.	n_{2c}	.	$n_{2.}$
.
.
C	n_{c1}	n_{c2}	.	.	.	n_{cc}	.	$n_{c.}$
n_{ij}	$n_{.1}$	$n_{.2}$.	.	.	$n_{.c}$.	$n_{..} (= n)$

The proposed method assumes that there are replications; that is, there is more than one pair of case-control subjects in various case-control subject treatment combinations. It is however, for simplicity, assumed that there are no interactions between combinations or that such interaction have been removed by appropriated data transformation.

2. The Proposed Method

Suppose the i^{th} pair of case-control subjects or patients is selected and members of the selected pair are each randomly assigned to treatments T_1 (Standard drug, Control) or T_2 (new drug, Case) and their various responses are recorded for $i = 1, 2, \dots, n_{ij}$ pairs; $n_{ij} = 1, 2, \dots, c$ responses and $n = \sum_{i=1}^c \sum_{j=1}^c n_{ij}$ is the total number of pairs of case-control subjects studied. As already noted above, the responses by members of each matched pairs of subjects or patients are classified into c-possible mutually exclusive categories or classes. Suppose that the responses by both case and control subjects and patients are quantitative and can assume all possible values on non negative real number line.

Now, let (y_{il}, y_{ij}) be the i^{th} pair of the scores or responses by case administered treatment T_2 (new drug) responding at the l^{th} response option or category and the corresponding control subject or patient administered treatment T_1 (standard drug) responding at the j^{th} response option or category for $i = 1, 2, \dots, n_{ij}$, $l, j = 1, 2, \dots, c$ response categories or options

$$\text{Let } y_{ilj} = y_{il} - y_{ij} \quad (1)$$

be the difference between the scores by case administered treatment T_2 (new drug) responding at the l^{th} response category and the corresponding control subject or patient administered treatment T_1 (standard drug) responding at the j^{th} response category by the i^{th} pair of case-control subjects or patients on the condition of interest for $i = 1, 2, \dots, n_{ij}$, $l, j = 1, 2, \dots, c$.

Also, let

$$y_i = \begin{cases} 1 & \text{if } y_{ilj} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

For $i = 1, 2, \dots, n$; $l, j = 1, 2, \dots, c$

We propose, develop and present the use of dummy variable regression models in the analysis of diagnostic screening test results with several case-control subject –response options or categories ($c \geq 2$) with replications in each case- control response categories or combination and for comparing the differences between proportion of cases and controls subjects responding under various response categories or options. (Fleiss (1981), Ryan (1997))

Consistent with the use of dummy variable in regression models (Boyle 1975), we represent the c-response categories or options by case **a** that is, “c” levels of T_2 (new drug) with $c - 1$ dummy variables of 1s and 0s and also the “c” response categories or options by control, “t”, that is, the “c” levels of T_1 (standard drug) by $c - 1$ dummy variables of 1s and 0s

Hence using the difference y_{ilj} between the scores by the i^{th} pair of case - control subjects or patients (Eqn 1) as the dependent or criterion variable, we may now define

$$x_{l,a} = \begin{cases} 1, & \text{if for the } i^{th} \text{ pair of case control subjects, the difference in scores,} \\ & y_{ilj} = y_i \text{ is such that case is in the } l^{th} \text{ response category for all the } i, j \\ & \text{response categories by the corresponding control subject in } T_1 \\ 0 & \text{otherwise} \end{cases}$$

Above expression is Equation (3). For $i = 1, 2, \dots, n$, and some $l = 1, 2, \dots, c - 1$ and all $j = 1, 2, \dots, c$. Also let

$$x_{ij,t} = \begin{cases} 1, & \text{if for the } i^{th} \text{ pair of case control subjects, the difference in scores,} \\ & y_{ilj} = y_i \text{ is such that case is in the } j^{th} \text{ response category for all the} \\ & l \text{ response categories by the corresponding case in treatment } T_2 \\ 0 & \text{otherwise} \end{cases}$$

Above expression is Equation (4). For $i = 1, 2, \dots, n_{lj}$ and some $j = 1, 2, \dots, c - 1$ and all $l = 1, 2, \dots, c$.

Using these dummy variables, set up a dummy variable multiple regression model of y_i regressing on the $2(c - 1)$ dummy variables of 1s and 0s of Eqn 3 and 4 expressing the dependence of the relative magnitude of the differences y_i between the scores of the i^{th} pair of case control subjects in the screening tests as

$$\begin{aligned} y_{ilj} = y_i = P(Y_{ilj} = 1) = P(1) = & \beta_0 + \beta_{1,a}x_{i1,a} \\ & + \beta_{2,a}x_{i2,a} + \dots + \beta_{c-1,a}x_{ic-1,a} + \beta_{1,t}x_{i1,t} \\ & + \beta_{2,t}x_{i2,t} + \dots + \beta_{c-1,t}x_{ic-1,t} \end{aligned} \quad (5)$$

Equation 5 is the probability that the i^{th} randomly selected case-control pair of responses on a given condition the score by case administered treatment T_2 (new drugs) is on the average higher (better, greater, less serious) than the score by the corresponding control subject administered treatment T_1 (standard drugs) or vice versa.

The odds of occurrence of this events, that is, the odds that for the i_{th} case-control treatment combination, the response by case administered treatment T_2 is higher (better, greater, less serious) than the response or score by the corresponding control subject administered treatment T_1 is

$$\begin{aligned} \Omega_{T_2,T_1} &= \frac{P_i(1)}{1 - P_i(1)} = \frac{P_i(1)}{P_i(0)} \\ &= \frac{\beta_0 + \sum_{i=1}^{c-1} \beta_{l,a}x_{il,a} + \sum_{j=1}^{c-1} \beta_{j,t}x_{ij,t} + \dots + b}{1 - \beta_0 + \sum_{i=1}^{c-1} \beta_{l,a}x_{il,a} + \sum_{j=1}^{c-1} \beta_{j,t}x_{ij,t}} \end{aligned} \quad (6)$$

for $l = 1, 2, \dots, n$

x_i^s are dummy variables of 1s and 0's, β 's are partial regression coefficients and e_i 's are error terms with $E(e_i) = 0$ for $i = 1, 2, \dots, n$

Now, the expected value or mean of $y_i = P_i(1)$; the relative difference in scores or responses by the i^{th} pair of case control subjects when the case responds at the l^{th} response category and control subject responds at the j^{th} response category or option is

$$\begin{aligned} E(y_{ilj}) &= E(y_i) = E(P_i(1)) \\ &= \beta_0 + \sum_{l=1}^{c-1} \beta_{l,a}x_{il,a} + \sum_{j=1}^{c-1} \beta_{j,t}x_{ij,t} \end{aligned} \quad (7)$$

Equation 5 may be alternatively expressed in matrix form as

$$y = P(1) = X\beta + e \quad (8)$$

where $y = P(1)$ is an $nx1$ column vector of 1s and 0s as specified in equation 1, X is an $nx(p + 1)$ design matrix of 1's and 0's where $p = (c - 1) + (c - 1) = 2(c - 1)$ dummy variables of 1's and 0's representing the $2(c - 1)$ response categories included in the regression model, β 's are partial regression coefficients and e is an $nx1$ column vector of error terms with $E(e) = 0$

The expected value of $y = P(1)$ is

$$E(y) = E(P(1)) = X \quad (9)$$

Use of the usual methods of least squares with either Equations 2 or 4 yields unbiased estimates of the regression parameter β as

$$\hat{\beta} = b = (X'X)^{-1}X'y = (X'X)^{-1}X'P(1) \quad (10)$$

where $(X'X)^{-1}$ is the matrix inverse of the non singular variance - covariance matrix $X'X$ with full column rank $[p + 1]$.

The corresponding fitted regression model is

$$\hat{y} = \hat{P}(1) = Xb \quad (11)$$

A null hypothesis that maybe of interest is that the regression model of either Equations 4 or 7 fits; that is, it is adequate representation of differences in scores or responses between pairs of case and control subjects in diagnostic screening tests. This is equivalent to the null hypothesis.

$$H_0: \beta = 0 \text{ versus } H_1: \beta \neq 0 \quad (12)$$

The null hypothesis H_0 of Equation 11 is tested using the usual F test presented in the traditional analysis of variance Table.

The null hypothesis H_0 of Equation 11 is rejected at the α level of significance if the calculated F ratio is greater than the tabulated or critical F ratio with "p" and "n - p - 1" degrees of freedom; that is, if

$$F \geq F_{1-\alpha; p, n-p-1}$$

Otherwise H_0 is accepted.

If the null hypothesis H_0 of Equation 11 is rejected, in which case not all the β_s are equal to zero, then one may proceed further to determine which of these regression coefficients or their combinations are statistically different from zero and hence, may have been responsible for the rejection of H_0 . In particular, one may wish to compare the responses of cases and controls in the paired case-control subject responses or scores.

Note that the expected difference in scores or responses between case and control subjects when the responses by the control subjects in the paired case-control subjects responses are held at constant level, that is when the mean response or score by case in all the pairs is of interest, is obtained by setting $x_{1l,t} = 1$ and $x_{1j,t} = 0$, for $l, j = 1, 2, \dots, c-1$ in Equation 5 yielding

$$E(y_{ilj;a}) = \beta_0 + \sum_{l=1}^{c-1} \beta_{lj,a} \quad (13)$$

Similarly, the expected difference in scores or responses between case and control subjects when the response by case in the paired case-control responses or scores are held at constant levels, that is when the mean response or score by the control subjects in all the pairs of interest is obtained by setting $x_{ij,t} = 1$ and $x_{il,a} = 0$ in Equation 5 yielding

$$E(y_{ilj;t}) = \beta_0 + \sum_{j=1}^{c-1} \beta_{j,t} \quad (14)$$

The difference between these expected values or mean responses is

$$\beta_{a,t} = \sum_{l=1}^{c-1} \beta_{l,a} - \sum_{j=1}^{c-1} \beta_{j,t} \quad (15)$$

whose sample estimate is

$$\hat{\beta}_{a,t} = b_{a,t} = \sum_{l=1}^{c-1} b_{l,a} - \sum_{j=1}^{c-1} b_{j,t} \quad (16)$$

Having rejected the initial null hypothesis H_0 of Equation 11, one may then wish to test the null hypothesis that $\beta_{a,t}$ is equal to some specified value say β_{00} that is

$$H_0: \beta_{a,t} = \beta_{00} \text{ versus } H_1: \beta_{a,t} \neq \beta_{00} \quad (17)$$

The null hypothesis H_0 of Equation 16 is tested using the “t” test statistic

$$t = \frac{(\sum_{l=1}^{c-1} b_{l,a} - \sum_{j=1}^{c-1} b_{j,t}) - \beta_{00}}{\sqrt{c'(X'X)^{-1}.cMSE}} \quad (18)$$

where $c' = (0, 1, 1, \dots, 1, -1, -1, \dots, -1)$ is a $1 \times (p+1)$ row vector of 0, $(c-1)1s$, and $(c-1-1s)$ with 1s occurring at $b_{l,a}$ and -1s occurring at $b_{j,t}$; $l, j = 1, 2, \dots, c-1$, and MSE is the error mean sum of squares obtained from the corresponding analysis of variance Table.

The null hypothesis of Equation 16 is rejected at the α level of significance if $t \geq t_{1-\frac{\alpha}{2}; n-p-1}$ otherwise H_0 is accepted.

Further research interest may also be in performing pairwise comparison of the effects scores by case and control subject at the l th response category for case and j th response category for control subjects have on the variation in the differences between scores or responses by the l th

pair of case-control subjects in the two response categories. This is equivalent to testing the null hypothesis

$$H_0: \beta_{l,a} = \beta_{j,t} \text{ versus } H_1: \beta_{l,a} \neq \beta_{j,t} \quad (19)$$

for $l, j = 1, 2, \dots, c-1$

The null hypothesis of Equation 18 is tested using the student “t” test statistic;

$$t = \frac{b_{l,a} - b_{j,t}}{\sqrt{\text{Var}(b_{l,a}) + \text{Var}(b_{j,t}) - 2\text{Cov}(b_{l,a}; b_{j,t})}} = \frac{b_{l,a} - b_{j,t}}{\sqrt{(c_{ll} + c_{jj} - 2c_{lj})\text{MSE}}} \quad (20)$$

C_{lj} is the element in the l^{th} row and j^{th} column of $(X'X)^{-1}$ for $l, j = 1, 2, \dots, c-1$

The null hypothesis H_0 of Equation 18 is rejected at the α level of significance if Eqn 19 is satisfied otherwise H_0 is accepted.

If in particular, the “c” response categories are ordered from the least serious (most serious) or the most serious (least serious) or if any of the response categories is treated or known to be in the best (least serious) condition or the worst (most serious) condition, then, the null hypothesis of Equation 18 would be equivalent to testing the null hypothesis that pairs of case-control subjects are equivalent in the same level of seriousness or condition in the test being conducted.

The test statistic of Equation 19 is used to test this null hypothesis; that is, the null hypothesis that case and control subjects are in the same condition, that is, do not differ in their responses or scores in the screening test.

Note that in practice, the data of interest are not usually in their summary form as frequencies as in Table 1 but as raw scores, that is, pairs of case-control subject responses or scores (y_{il}, y_{ij}) for $i = 1, 2, \dots, n_{lj}$; $l, j = 1, 2, \dots, c-1$, in these cases, the frequencies if required may nevertheless be obtained from the design matrix X used for the analysis. In other words, the cell frequencies n_{lj} are obtained from the columns X as;

$$n_{lj} = x'_{l,a}; x_{j,t} \text{ for } l, j = 1, 2, \dots, c-1. \quad (21)$$

Also, $n_{l,c}$ and $n_{c,j}$ are obtained as the scores in the omitted categories for case T_2 and control T_1 response categories respectively.

$$\begin{aligned} n_{lc} &= x'_0 \cdot x_{l,a} - \sum_{j=1}^{c-1} n_{lj}; n_{cj} \\ &= x'_0 \cdot x_{j,t} - \sum_{l=1}^{c-1} n_{lj}; \end{aligned} \quad (22)$$

for $l, j = 1, 2, \dots, c-1$; and x_0

is an $n \times 1$ column vector of 1s.

The number of pairs of case-control responses or scores n_{cc} in the $(c, c)^{\text{th}}$ case-control response category, that is the number of scores common to the “omitted categories” for case and control subjects in the design matrix X of the regression model is obtained as the number of zeros common to all the $c-1$ “included categories” for case and control subjects respectively, or equivalently as

$$n_{c,t} = n - \sum_{l=1}^{c-1} \sum_{j=1}^{c-1} n_{lj} \quad (23)$$

with these frequencies, one may now apply an alternative generalized method based on only frequencies (Oyeka et al, 2013) to the data for comparative purposes.

2.1. Illustrative Example

We, now use data on matched subjects of $n = 54$ matched patients from a controlled comparative clinical trial who manifest four possible responses to illustrate the proposed method. Suppose the data in Table 2 are obtained by assigning a standard treatment T_1 regarded as control drug and a new treatment T_2 regarded as case at random to members of each pair of a random sample of 54 pairs of case and control malaria patients matched on age, gender and body weight used in controlled clinical trials to compare the effectiveness of two malaria drugs.

Using these y_{ilj} values of Table 2 with Equations 1-4, we obtain the design matrix X shown in Table 3 for use in the regression model of Equation 9.

Using the design matrix X of Table 3 in Equation 6, we

obtain the fitted dummy variable regression model of the difference in case-control subject response scores as function of the seriousness of it or levels of these responses as

$$y_{ilj} = -0.116 + 0.130x_{1j,a} + 0.243x_{2j,a} - 0.112x_{3j,a} + 0.016x_{1j,t} - 0.152x_{2j,t} - 0.128x_{3j,t} \\ (R^2 = 0.050, F = 0.410, P - value = 0.869) \quad (23)$$

The fitted regression model of Equation 23 with an $R^2 = 0.050$ implies that the independent variables can explain only about 5 percent of the total variation in case and control subjects in their response scores, suggesting that its regression coefficients are all probably not different from zero.

With these conclusions, further analysis would ordinarily become unnecessary, however, for illustrative purposes only, one could assume that not all the regression coefficients are zero thereby enabling some comparisons of these parameters if so desire.

Table 2. Sample Data of Response Scores by Malaria patients in controlled Clinical Trials with Replications

		Control (Standard Drug T_1)							
Case (New Drug T_2)		Improved		Fair		Serious		Critical	
Improved		(1.17, 1.33)(1.80,1.43)		(1.60,1.62)(1.47,1.47)		(1.62,1.50)(0.85, 1.21)		(0.70, 1.51)(0.88,1.09)	
		-0.16	0.37	-0.02	0.00	0.05	-0.43	-0.61	-0.21
		(1.97, 1.36)(0.82, 1.41)		(1.71, 1.18)(0.97, 0.85)		(0.83,1.19)		(1.29,0.93)	
		0.61	-0.59	0.53	0.12	-0.36		0.36	
Fair				(0.83, 1.39)(1.46, 1.56)					
				-0.56	-0.10				
		(1.28, 1.26)(1.30, 0.99)		(0.85, 1.24)(1.12, 1.23)		(1.33, 1.37)(1.02, 1.33)		(0.90,0.93)(1.62, 0.55)	
		0.02	0.31	-0.31	-0.11	-0.04	-0.31	, 0.03	0.74
Serous		(1.27, 1.47)(1.24, 1.09)		(1.19, 1.01)(1.12, 0.21)		(0.80, 1.36)(1.60, 0.74)		(1.55, 1.09)	
		-0.20	0.15	-0.14	-0.91	-0.56	0.36	0.46	
				(1.92, 1.26)					
				0.66					
Critical		(1.28, 0.89)(0.84, 1.64)		(1.15, 2.02)(1.37, 0.83)		(1.63, 1.62)(0.97, 0.86)		(0.82, 0.73)(1.54, 0.89)	
		0.39	-0.80	-0.87	0.54	0.01	0.11	0.09	0.67
		(1.80, 1.32)		(1.32, 1.57) (1.57, 1.85)		(0.94, 2.01)			
				0.48		-0.35	-0.28	-1.07	
		(0.93, 0.74)(0.67, 0.89)		(1.05, 2.01)(1.43, 1.93)		(1.62, 1.70)(0.85, 1.54)		(1.51, 1.07)(1.27, 1.99)	
		0.19	-0.22	-0.96	-0.50	-0.08	-0.99	0.44	-0.72
		(1.53, 0.90)				(1.62, 0.93)			
		0.65				0.69			

Table 3. Design Matrix X for the data of Table 2

	Diff y_{ilj}	0 x_{i0}	1 $x_{i1,a}$	2 $x_{i2,a}$	3 $x_{i3,a}$	4 $x_{i4,a}$	5 $x_{i5,a}$	6 $x_{i6,a}$
1	-016	1	1	0	0	1	0	0
2	0.37	1	1	0	0	1	0	0
3	0.61	1	1	0	0	1	0	0
4	-0.59	1	1	0	0	1	0	0
5	-0.02	1	1	0	0	0	1	0
6	0.00	1	1	0	0	0	1	0
7	0.53	1	1	0	0	0	1	0
8	0.12	1	1	0	0	0	1	0
9	-0.56	1	1	0	0	0	1	0
10	-0.10	1	1	0	0	0	1	0
11	0.05	1	1	0	0	0	0	1
12	-0.43	1	1	0	0	0	0	1
13	-0.36	1	1	0	0	0	0	1
14	-0.61	1	1	0	0	0	0	0
15	-0.21	1	1	0	0	0	0	0
16	0.36	1	1	0	0	0	0	0
17	0.02	1	0	1	0	1	0	0
18	0.31	1	0	1	0	1	0	0
19	-0.20	1	0	1	0	1	0	0
20	0.15	1	0	1	0	1	0	0
21	-0.39	1	0	1	0	0	1	0
22	-0.11	1	0	1	0	0	1	0
23	0.14	1	0	1	0	0	1	0
24	-0.91	1	0	1	0	0	1	0
25	0.66	1	0	1	0	0	1	0
26	-0.04	1	0	1	0	0	0	1
27	0.31	1	0	1	0	0	0	1
28	-0.56	1	0	1	0	0	0	1
29	0.86	1	0	1	0	0	0	1
30	0.03	1	0	1	0	0	0	0
31	0.74	1	0	1	0	0	0	0
32	0.46	1	0	1	0	0	0	0
33	0.39	1	0	0	1	1	0	0
34	-0.80	1	0	0	1	1	0	0
35	-0.87	1	0	0	1	0	1	0
36	0.54	1	0	0	1	0	1	0
37	0.48	1	0	0	1	0	1	0
38	0.01	1	0	0	1	0	0	0
39	0.11	1	0	0	1	0	0	0
40	-0.25	1	0	0	1	0	0	0
41	-0.28	1	0	0	1	0	0	0
42	0.09	1	0	0	1	0	0	0
43	0.67	1	0	0	1	0	0	0
44	-1.07	1	0	0	1	0	0	0
45	0.19	1	0	0	0	1	0	0
46	-0.22	1	0	0	0	1	0	0
47	0.35	1	0	0	0	1	0	0
48	-0.96	1	0	0	0	0	1	0
49	-0.50	1	0	0	1	0	1	0
50	-0.08	1	0	0	0	0	0	1
51	-0.99	1	0	0	1	0	0	1
52	0.69	1	0	0	1	0	0	1
53	0.44	1	0	0	0	0	0	0
54	-0.72	1	0	0	1	0	0	0

3. Summary and Conclusions

From the result of the analysis, $R^2 = 0.050$ implies that independent variables can only explain 5 percent of the total variation in case and control subjects in their response scores, suggesting that its regression coefficients are all probably not different from zero (insignificant). Also, the p-value of 0.869 indicates that the model is insignificant. The illustrative example shows that the proposed method is adequate for modeling and determination of significance/insignificance models.

REFERENCES

- [1] V.P. Backpar (1966): A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statistical Association*, 1966, 61, 228-235.
- [2] C. R. Boyles (1975). A problem based learning approach to teaching biostatistics. *Journal of Statistics Education* vol.7, n.1.
- [3] JL Fleiss (1981). *Statistical methods for rates and proportions (second ed.)* New York: Wiley, 1981.
- [4] J. F. Grizzle., C. F. Starmer and G. G. Koch. (1969): Analysis of categorical data by linear models *Biometrics*.
- [5] C. Ireland, H.H. Ku, and S. Kullback (1969): Symmetry and marginal homogeneity of an $r \times r$ contingency table. *J. Am. Stat. Assoc.* 64, 1323-1341.
- [6] A.E Maxwell (1970): Comparing the classification of subjects by two independent judges. *British Journal of Psychiatry*, 1970, 116, 651-655.
- [7] I.C.A. Oyeka, and, U.M. Okeh. (2013) Estimation of subject specific index of relative performance in "K" samples. *American Journal of theoretical and applied statistics* vol. 2 pg 154-165
- [8] T.. Ryan, (1997), *Modern Regression Methods*, John Wiley.
- [9] T. P. Ryan, and N. C. Schwertman, (1997). "Optimal limits for attributes control charts", *Journal of Quality Technology*, 29 (1), 86-98.
- [10] A.A. Staurt. (1955): A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, 1955, 42, 412-416.