IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 11, Issue 1 Ver. III (Jan - Feb. 2015), PP 12-15
www.iosrjournals.org

Bipolar Disorder Investigation Using Modified Logistic Ridge Estimator

Ogoke U. P.,[1] Nduka E.C. ,[1] Nja M.E. [2]
1.Department of Mathematics and Statistics, University of Port Harcourt, Rivers State, Nigeria.
2. Department of Mathematics, Federal University Lafia, Nigeria.

Abstract: *Bipolar affective disorder is a mental illness classified as mood disorder. Many people with bipolar disorder still go about their normal day-to-day schedule of duties. This is quite dangerous for the civil society. As the severity increases, affected persons become more erratic, more violent and often make poor judgments.In this paper, factors other than the traditional, genetic, environmental and neurochemical have been investigated among one hundred and nine teaching and non-teaching staff of a university. These are sex, age, occupation and body mass index (BMI). Using the Modified Logistic Ridge estimator, sex, age and body mass index have been found to contribute significantly to bipolar disorder. Both the standard error test and the odd ratios were used to determine the significance or otherwise of the factors. Men are more predisposed, older people ($\geq 40 years$) are more predisposed and Body mass index is positively correlated to bipolar disorder. The model can be used to determine the probability of involvement in each of the eight sub populations. Intervention should focus on men, people aged 40 and above and people with a higher body mass index.*
Keywords: *Bipolar Disorder, Collinearity,Logistic Estimator,Logistic Ridge Estimator, Modified Logistic Ridge Estimator, Odd ratios.*

I. Introduction

Bipolar disorder (BD) is a complex, chronic mood disorder involving repeated episodes of depression and mania/hypomania (American Psychiatric Association 2001). In addition to multiple relapses of mood episodes,individuals living with BD also experience substantial symptoms between episodes (Schaffer et al. 2006;Benazzi2004; Paykel et al. 2006).

Bipolar disorder is marked by extreme mood swings from highs to lows. These episodes can last for hours, days, weeks or months. The mood swings may even become mixed, so you might feel like crying over something upbeat.According to the National Institute of mental Health & other authorities, B.D may include these warning symptoms; feeling overly happy or optimistic for long stretches of time, feeling easily agitated-some describe it as feeling jump or "twitchy", talking fast, restlessness or impulsiveness, impaired judgment, over confidence, and engaging in risky behavior such as having impulsive sex, gambling ones savings or going on big spending sprees.

Bipolar disorder is not curable but an effective treatment or management carried out on a patient over a long period of time will be of immense benefit. Such treatments will help patients even in the most severest forms of their mood swings and related symptoms (Sachs etal. 2000; Hurley et al. 2000).

Brook etal (2006) used regression analysis to compare the cost of employees with BD with that of other employee cohorts, to assess cost differences. They built separate regression models for each dependent variable considered. Logistic regression was also used to model the medical costs and predict likelihood of having any medical costs during the year. Out of the four cohorts considered,comparisons of cohorts 1,2,3 and 4 yielded significant differences except between cohort 1 and 3 dealing with sick leave cost.

Collinearity

In both General linear models (GLMs) and Generalized linear models (GLMs), collinearity exists among two or more independent variables which are highly correlated (Mason, 1987). The effects of this are to produce regression based parameter estimates with inflated variances.

Gunst(1984) defines collinearity to exist among the columns of$X(X_1, X_{2,}, \ldots, X_p)$

if for suitably predetermined $\varepsilon_n > 0$, there exists constants$C_1, C_2, \ldots\ , C_p$ not all zero such that$C_1X_1 + C_2X_2 +, \ldots\ , +C_pX_p = S$, with $\|S\| < \varepsilon_n \|C\|$.

If the goal is simply to predict Y from a set of X variables, then multicollinearity is not a problem. The prediction will still be accurate and the overall R^2 (or adjusted R^2) quantities how well the model predicts Y values.Motilskey (2002),Weissfield&Sereika (1991) obtained a collinearity diagnosis for GLMs by performing the singular value decomposition on the scaled observed information matrix.

The collinearity problem can be addressed by the following existing methods: Variable Selection, Ridge Regression, and Principal Component Analysis etc.

Ridge Regression

This procedure is used in combating collinearity in GLM using the correlation matrix, standardized regression coefficient and by introducing the biasing constant C, into the normal equations of the standardized regression model. The ridge regression procedure can be stated using the following normal equation Tikhonov etal (1998),

$$\gamma_{XY}\hat{\beta} = \gamma_{XX} \quad ,$$

where$\gamma_{XY}$is the matrix correlation of Y with X and γ_{XY} is the matrix correlation of X. Adding the biasing constantC to the normal equation, we

$$(\gamma_{XX} + CI)\bar{\beta}^{R} = (\gamma_{XX} + CI)^{-1}\gamma_{XX}$$

Several successive values of C are tried until the regression coefficient becomes stable.By Ordinary Least Squares estimation, we consider the system $X\beta = Y$

II. Method

The Logistic Ridge Estimator

The modified logistic Ridge estimator derives from the Logistic Ridge estimator as an extension of the logistic estimator. In logistic Ridge regression, a biasing constant is introduced into the information matrix in order to overcome the problem of collinearity. Where collinearity exists among explanatory variables, the ordinary logistic regression becomes inadequate because of singularity problems.

The logistic regression model is given by the log of an odd or logit of a response probability as follows

$$\text{Logit}(\mu) = \text{Log}\left(\frac{\mu}{1-\mu}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k \qquad (1)$$

where$\mu = \left\{1 + \exp\left[-(\hat{\beta}_0 + \sum_{k=1}^{t} \beta_k x_k)\right]\right\}^{-1}$.The expression$\frac{\mu}{1-\mu}$ is called an odd of a favorable outcome and it is expressed as

$$\frac{\mu}{1-\mu} = \exp\left\{\beta_0 + \sum_{k=1}^{t} \beta_k x_k\right\}$$

Let μ_{hijk} denote the response probability for the h^{th} sex status, i^{th}age category, j^{th} occupational status, and k^{th} Body Mass Index (BMI).The Logistic model is then written as

$$\log\left(\frac{\mu_{hijk}}{1-\mu_{hijk}}\right) = \hat{\beta}_0 + \sum_{k=1}^{t} \beta_k X_{hijk}$$

Maximum Likelihood (McCullagh and Nelder (1972), Der and Everitt (2009)) is used to estimate the parameters of the logistic model in equation (1).The log likelihood function for the logistic model is given as

$$L(\beta;y) = \sum_i y_i \log[\mu(\beta' x_i)] + (1 - y_i)\log[1 - \mu(\beta' x_i)],$$

where$y' = [y_1, y_2, \ldots, y_n]$ are the n observed values of the dichotomous response variable.However, because of the singularity of the ordinary logistic model as earlier mentioned, the logistic Ridge regression is used to estimate this parameter values.

In this work, collinearity has been established among the explanatory variables under study. It is to be observed that the variables under investigation are both categorical and continuous. A test for collinearity using standard errors of parameter estimates and condition numbers reveal the existence of collinearity among the explanatory variables. Running the analysis using both the logistic estimator and the logistic Ridge estimator separately, reveals that the Ridge estimator has smaller standard errors. The logistic Ridge estimator (Tikhonov et al (1998)) is given as

$$\hat{\beta} = (X'WX + \theta I)^{-1}X'Z, \qquad (2)$$

Where,θ is the Tikhonov constant, Z is the adjusted dependent variable, W is a weight matrix, and X is the designed matrix.

The initial value of this constant is normally intelligently guessed. This constant can also be generalized as in the case of Generalized Logistic Ridge regression. The components of W and Z are given respectively as

$$w_i = m_i \mu_{i(1-\mu_i)}$$

$$z_i = \eta_i + (y - \mu_i)\frac{1}{\mu_i(1 - \mu_i)}$$

Variance inflation of parameter estimates is the problem introduced by collinearity. To address this problem, the logistic ridge regression is introduced. However, the ridge regression estimator is biased. To solve the problem of bias, we modify the logistic ridge estimator by exponentiating the response function. This yields the following modified logistic Ridge estimator (Ogoke, et al (2013)).

$$\hat{\beta} = \left(X'W_{\sqrt{1+\gamma}}X + CI\right)^{-1} X'W_{\sqrt{1+\gamma}}Z_{\sqrt{1+\gamma}} \qquad (3)$$

where elements of $W_{\sqrt{1+\gamma}}$ and $Z_{\sqrt{1+\gamma}}$ are respectively defined as

$$W_{i\sqrt{1+\gamma}} = m_i \mu_i^{\sqrt{1+\gamma}}(1-\mu_i^{\sqrt{1+\gamma}})$$

$$Z_i = \eta_i + \left(\frac{y_i}{m_i} - \mu_i^{\sqrt{1+\gamma}}\right)\frac{1}{\mu_i^{\sqrt{1+\gamma}}\left(1-\mu_i^{\sqrt{1+\gamma}}\right)}$$

$\mu_i^{\sqrt{1+\gamma}}$ is a modified response function(n is number of subpopulation, i, ..., n, and $-1 < \gamma \leq 1$).

Using the MATLAB package, we developed a programme to execute the algorithm that yields the update (3). Our results are compared with the results of the Logistic Estimator and those of the Logistic Ridge Estimator. The results show that the variance of the proposed Modified Logistic Ridge Estimator is smaller than that of the logistic ridge Estimator which in turn is smaller than that of the Logistic Estimator. A collinearity test using condition numbers was used to ascertain the existence of collinearity before deciding on the proposed Modified Logistic Ridge estimator.

The Modified Logistic Ridge estimator is superior to the existing Logistic Ridge estimator in the sense that the later estimator is a special case of the former. The choice of γ in the update provides a flexibility that serves as an advantage for this estimator.

III. Results

The results now follow

Table 1: Logistic Estimator

Parameter Estimates	Standard Errors
Constant ($\hat{\beta}_0$)= -2.8580	SE($\hat{\beta}_0$)=1.860349
Sex($\hat{\beta}_1$)= 0.5111	SE($\hat{\beta}_1$)= 0.240593
Age($\hat{\beta}_2$)= 0.6358	SE($\hat{\beta}_2$)= 0.222703
Occupation($\hat{\beta}_3$)=-0.0486	SE($\hat{\beta}_3$)= 0.163360
BMI($\hat{\beta}_4$)=0.0839	SE($\hat{\beta}_4$)=0.000361

Table 2: Logistic Ridge Estimator

Parameter Estimates	Standard Errors
$\hat{\beta}_0 = -2.214$	SE($\hat{\beta}_0$) = 4.053
$\hat{\beta}_1$(Sex) $= -0.535$	SE($\hat{\beta}_1$) = 0.460
$\hat{\beta}_2$(Age) $= -0.633$	SE($\hat{\beta}_2$) = 0.422
$\hat{\beta}_3$(Occupation) = 0.056	SE($\hat{\beta}_3$) = 0.438
$\hat{\beta}_4$(BMI) = 0.101	SE($\hat{\beta}_4$) = 0.150

Table 3: Modified Logistic Ridge Estimator

Parameter Estimates	Standard Errors
$\hat{\beta}_0 = -3.186994$	SE($\hat{\beta}_0$) = 3.45611
$\hat{\beta}_1$ = 0.538102	SE($\hat{\beta}_1$) = 0.05787
$\hat{\beta}_2$ = 0.677427	SE($\hat{\beta}_2$) = 0.04957
$\hat{\beta}_3 = -0.046637$	SE($\hat{\beta}_3$) = 0.02668
$\hat{\beta}_4$ = 0.013	SE($\hat{\beta}_4$) = 0.00001

IV. Discussions And Conclusions

From tables 1, 2, and 3 respectively, it can be seen that the modified Ridge estimator is superior to the existing ones. The response probabilities are used to estimate individual probabilities for the various subgroups. The following response probabilities were obtained for the 8 sub-groups:

μ_1=0.3406, μ_2= 0.3773, μ_3 = 0.4905, μ_4 = 0.5115, μ_1=0.4322,μ_6=0.4492,μ_7=0.6943, μ_8 = 0.5875.

The probability of bipolar tendency is highest (0.69) for non-teaching males who are not less than 40 years of age and who have a BMI 30.16(the highest of BMI). This is followed by teaching males who are 40 and above with BMI 25.18.

From this result, we can conclude that older men are more predisposed to bipolar disorder than their female or younger counterparts.A further analysis uses odd ratios. This shows that the odd ratio of bipolar disorder for females versus males at teaching level$\geq$ 40years is 0.73 or 73:100 against the males. That of females versus males at teaching level but <40years is 74:100.Again, the odds are against males. This again supports the fact of the response probability stated above that men are more predisposed to bipolar disorder than females.

Effective screening programs along with early identification and intervention are increasingly important in today's economic climate that places enormous emphasis health care costs and optimizing employee productivity. Accurate and timely recognition of BD has the potential to reduce medical costs and indirect costs due to work loss.

References

[1]. American Psychiatric Association (2001). Diagnostic and Statistical Manual of Mental Disorders (4thed.).Washington, DC: TeRevision (DSM-IV-TR).
[2]. Benazzi, F. (2004).Inter-episode mood liability in mood disorders: residual symptom or natural course of illness? Psychiatry ClinNeurosci.58(5):480-6.
[3]. Brook, A.R et al (2006). Incurring Greater Health care Costs; Risk Stratification of Employees with Bipolar Disorder.Journal of Clinical Psychiatry, Vol.8 (1).
[4]. Der, G.&Everitt, S. (2009).A Handbook of Analysis usingSAS (3rd edition).Boca Raton: CRC Press,Champion and Hall.
[5]. Gnust, R.F (1984).Toward a Balanced Assessment of Collinearity Diagnostics. American Statistician.Vol.38, p. 79-82.
[6]. Huxley NA, Parikh SV, Baldessarini RJ.(2000) Effectiveness of psychosocial treatments in bipolar disorder: state of the evidence. Harv Rev sychiatry.8(3):126–140.
[7]. Mason, G. (1987). Coping with Collinearity University of Manitoba Research Ltd. The Canadian Journal of Program Evaluation
[8]. McCullagh, P.&Nelder, J.A. (1992).Generalized Linear Models. Chapman and Hall, Madras.
[9]. Mental Health: A Report of the Surgeon General. U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health. 1999
[10]. Motalskey, H. (2002).Multicollinearity in multiple Regression President, Graph pad Software (hmotulsky @ graphpad.com).
[11]. Ogoke U.P, Nduka E.C, &NjaM.E (2013). A New Logistic Ridge Regression Estimator using ExponentiatedResponse Function.Journal of Statistical and Econometric Method.Vol. 2 (4), p. 161-171.
[12]. Paykel E.S, Abbott R, Morriss R, Hayhurst H, Scott J.(2006), Sub-syndromal and syndromal symptoms in the longitudinal course of bipolar disorder. Br J Psychiatry. 189:118-23
[13]. Sachs GS, Printz DJ, Kahn DA, Carpenter D, Docherty JP. (2000): Medication Treatment of Bipolar Disorder, The Expert Consensus) Guideline Series Postgrad Med. Apr;Spec No.:1–104.
[14]. Sachs G.S, Thase,M.E(2000). Bipolar disorder therapeutics:treatment. BiolPsychiatry.48 (6):573–581.
[15]. Tikhonov, A.N, Leonov, A.S, Yagola, A.G (1998): Non linear ill-posed problems, vol. 1, vol. 2 Chapman & Hall.
[16]. Weissfield&Sereika (1991): A Multicollinearity Diagnostic for Generalized Linear Models.