Comparison between Robust and Classical Analysis in Bivariate Logistic for Medical Data

Fadhil Abdul AA*
Department of Statistics, Al-Furat Al-Awast Technical University, Iraq

Abstract

Representing medical data and biological important part in experiments are concerned with Human life, the primary objective of this research is to use the statistical optimization method analysis for the data and knowledge of the important factors affecting the variables of the study (liver fat, liver size), where the variables are interconnected there is a need for statistical method to examine the degree of their relationship, we used bivariate logistic.

To achieve the of the research on the field study will be done in Al-Sadr medical city in the province of Najaf by taking a sample of 150 people auditors diabetes and liver disease center, from the statistical analysis results we observed the degree of diagnosis model in both method are good, and also we monitored that impact factors in responses (liver fat, liver size) and some comment as multivariate logistic in the Future.

Keywords: Bivariate logistic; Robust analysis; Logistic regression; Binary regression

Aim of Study

This research aims to review the method of bivariate logistic distribution in order to study and analysis effecting factors on the response variables (the degree of liver fatty and increase of liver size in Human beings) using the data of medical tests to compare classical and robust analysis when some values are outlier in the sample.

Introduction

In the medical and biological field studies, the experimentations are often related to the nature of the response adopted for non-continuous variable data (variables), but is the occurrence/non-occurrence of the score after taking a certain treatment (having effect or non-effect), regardless of the nature of the variables of the study at the left side of the General model equation \( y = x\beta + \varepsilon \), whether continuous, discrete or categorical.

Where the method of analysis depending on the type of data at the right side (\( y \)). If the binary response (0,1 depends on binary logit the method while ordinal response if rank \( (y) \)). The degree of healing of an illness or the degree of incidence of a particular disease is formulated into (generalized) ordinal logit/probit regressions, but if the response takes symbolic letters (A, B, ... ) can rely on logit (probit) multinomial.

From a historical view Barry W [1] studied binary data analysis with bivariate response under the influence of some independent variables and assuming a correlation between paired observations a disturbance depending on the Logistic regression model to estimate parameters the results were obtained are efficient compared with MLE to other researchers.

Kimberlee G, et al. [2] studied the analysis of correlated binary outcomes using Multivariate Logistic Regression for the case of two outcomes, a form of the cumulative bivariate logistic distribution proposed by Gumbel is used to characterize their joint probabilities in terms of logistic marginal probabilities and the correlation coefficient of the responses. They applied this technique in two different situations. When the correlation among responses is not significantly/and is significantly from zero.

In 1997 Sean M., David B [3] studied Bayesian analyses of multivariate binary (categorical) outcomes for \( \logit(Pr(y_i=1|x_j)) = x_j\beta \); where \( \beta \) is a vector of unknown regression coefficients with prior Normal distribution \( \pi(\beta)=N(\beta_0, \Sigma) \).

Thomas Y [4] studied Bivariate Binomial Responses by vgam family Functions \( B= \) breathlessness, \( W= \) wheeze; \( B= (i, W= (j; i,j=0,1) \), but Hun, M. in 2009 studied the Regression Models for binary dependent variables and the analysis data by Using Stata, SAS etc., He used data from clinicians and practitioners simulation study two responses (trust 1 respondent, 0 otherwise, www internet used 1 respondent, 0 otherwise) and five independents variables.

In 2014 Tabatabai, MA [5] and others had studied methods for robust Logistic and probit compared with MLE when there are outlier values, they have been rely on real data and simulation experience for \( (x_i=1,2) \) as independent variables, they proved robust method is efficient. but they didn't apply bivariate logistic of response.

Theory

Logistic distribution

In this part of paper, we display some basic concepts of application distribution Logistic experiments in which the data to the variable appears to stop responding (adopted) is continuous or binary data such as nominal or countable (classified), which does not require a well-known hypotheses linear regression model and there should be no outliers in the data, logistic regression assumes linearity of independent variables and log odds. Only it requires quite large sample sizes, because of maximum likelihood estimate require classification according to the

*Corresponding author: Fadhil Abdul Abbas Al-Abidy, Department of Statistics, Al-Furat Al-Awast Technical University, General Street-Player city, Najaf, Iraq; Tel: +9647801376429; E-mail: Abidy_fadhil@yahoo.com

Received October 06, 2015; Accepted October 26, 2015; Published November 02, 2015

Citation: Fadhil Abdul AA (2015) Comparison between Robust and Classical Analysis in Bivariate Logistic for Medical Data. J Biom Biostat 6: 262. doi:10.4172/2155-6180.1000262

Copyright: © 2015 Fadhil Abdul AA. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
number of response variables (the number of dependent variables) classified to [6,7]:

**Binary regression model**

This model depends on the following equation:

\[ y_i = \pi(x_i; \beta) + \varepsilon_i \quad ; i = 1,2,\ldots,n \]

Where \( \varepsilon_i \) are independent \( \forall i = 1,2,\ldots,n \)

\[ y_i \sim IBr \text{with men} = \pi(x_i; \beta) \quad \text{and var} = \pi(x_i; \beta)(1 - \pi(x_i; \beta)) \]

;\( i = 1,2,\ldots,n \)

a binary response variable \( (Y=1 \text{ or } Y=0) \) is associated a set of explanatory variables, as from the following functions:

\[ \pi = \text{prob}\{y = \text{outcomes of itersted} \mid X = x\} \]

\[ \logit\{p(y=1)\} = \theta(x; \hat{\alpha}) \]

is call simple logistic regression

**Multiple logistic model**

\[ \logit(\pi(x)) = \ln\left(\frac{p(y=1)}{1 - p(y=1)}\right) \]

\[ = \beta_0 + \beta_1x_1 + \beta_2x_2 + \ldots + \beta_px_p + \varepsilon \]

After a series of mathematical operations we get the following formula of binary multiple regression:

\[ \pi = \frac{\exp(\sum_{j=1}^{p} \beta_j x_j)}{1 + \exp(\sum_{j=1}^{p} \beta_j x_j)} ; x_0 = 1 \]

The data were analyzed according to the following cases:

**Classical method:** Do not use Robust analysis we got the following results

Through the results shown in the Table 1 above it is clear to display the following:

The test value (Likelihood Ratio LR) show an appropriate model

\[ N_2\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix} \mid \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right] = \mathcal{O}(\varepsilon, \varepsilon, \rho) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp \left[ \frac{-1}{2(1 - \rho^2)} (\varepsilon_1^2 + \varepsilon_2^2 - 2\rho\varepsilon_1\varepsilon_2) \right] \]
Table 1: Result analysis of bivariate logistic model by classical method.

| Variable   | Coef.    | Std. Err | z     | P>|z| | [95% Conf. Interval] |
|------------|----------|----------|-------|-----|---------------------|
| Fatty_type |          |          |       |     |                     |
| Age        | -0.0138739 | 0.0126286 | -1.09 | 0.267 | -0.036826 | 0.0018778 |
| Gender     | 0.0034449 | 0.2610207 | 0.01 | 0.998 | -0.508146 | 0.515036  |
| BMI        | 0.0701801 | 0.269094 | 2.61 | 0.009 | 0.174386  | 0.122916  |
| HbAc       | -0.1545345 | 0.053633 | -1.81 | 0.070 | -0.321853 | 0.012766  |
| Cho        | -0.0001859 | 0.0025515 | -0.07 | 0.942 | -0.005187 | 0.004815  |
| TG         | 0.0008252 | 0.0014075 | 0.59 | 0.558 | -0.001933 | 0.003583  |
| Pressure   | 0.0074953 | 0.0076659 | 0.97 | 0.331 | -0.076575 | 0.224843  |
| Goe        | 0.0296733 | 0.0062184 | 2.15 | 0.031 | 0.11168  | 0.094098  |
| Gpe        | 0.0184007 | 0.0386218 | 0.48 | 0.634 | -0.054972 | 0.030822  |
| ALI<       | 0.0019094 | 0.0045473 | 0.42 | 0.675 | -0.007003 | 0.001822  |
| cons       | -2.474073 | 1.517789 | -1.63 | 0.103 | -5.448866 | 0.503796  |
| Liver_type |          |          |       |     |                     |
| Age        | 0.0249884 | 0.0125168 | 2.00 | 0.046 | 0.000459  | 0.495209  |
| Gender     | 0.1890935 | 0.0448802 | 0.77 | 0.440 | -0.290693 | 0.694094  |
| BMI        | 0.0045514 | 0.0225854 | -0.20 | 0.840 | -0.048818 | 0.037913  |
| HbAc       | -0.0827778 | 0.0824425 | -1.07 | 0.284 | -0.249862 | 0.073306  |
| Cho        | 0.0070667 | 0.0026851 | 2.48 | 0.013 | 0.001479 | 0.021655  |
| TG         | 0.0019279 | 0.0013507 | -0.96 | 0.337 | -0.003945 | 0.001349  |
| Pressure   | 0.0059725 | 0.00675 | 0.88 | 0.376 | -0.007257 | 0.001922  |
| Goe        | 0.1645504 | 0.0582545 | 2.82 | 0.005 | 0.003737 | 0.278727  |
| Gpe        | 0.0513203 | 0.0041128 | 12.06 | 0.016 | 0.001338 | 0.003583  |
| ALI<       | -0.0023338 | 0.0041128 | 12.06 | 0.016 | 0.001338 | 0.003583  |
| cons       | -3.352374 | 1.492694 | -2.25 | 0.025 | -0.00206 | 0.001338 |
| jathro     | 0.12393 | 0.1580029 | 0.78 | 0.433 | -0.15875 | 0.43386  |
| rho        | 0.1232995 | 0.1556008 | 1.23 | 0.218 | -0.00206 | 0.001338 |

Table 2: Result analysis of bivariate logistic model by Robust method.

| Variable   | Coef.    | Std. Err | z     | P>|z| | [95% Conf. Interval] |
|------------|----------|----------|-------|-----|---------------------|
| Fatty_type |          |          |       |     |                     |
| Age        | -0.0138739 | 0.0135203 | -1.03 | 0.305 | -0.040373 | 0.0126253 |
| Gender     | 0.0034449 | 0.241705 | 0.01 | 0.989 | -0.470288 | 0.477176  |
| BMI        | 0.0701801 | 0.266763 | 2.62 | 0.009 | 0.0177254 | 0.1262349 |
| HbAc       | -0.1545345 | 0.0847402 | -1.82 | 0.068 | -0.320631 | 0.115443  |
| Cho        | -0.0001859 | 0.0024457 | -0.08 | 0.939 | -0.004879 | 0.0040676 |
| TG         | 0.0008252 | 0.001472 | 0.56 | 0.575 | -0.00206 | 0.001338 |
| Pressure   | 0.0074953 | 0.0073978 | 1.00 | 0.318 | -0.007257 | 0.001338 |
| Goe        | 0.0184008 | 0.0386689 | 0.48 | 0.634 | -0.057398 | 0.0491904 |
| Gpe        | 0.0019094 | 0.0044858 | 0.42 | 0.671 | -0.006902 | 0.001822  |
| ALI<       | -2.474073 | 1.703948 | -1.45 | 0.147 | -5.81375 | 0.8656052 |

Figure 1: Illustrative screen.

The results in Robust analysis are not different without Robust, but not significant (p-value=0.458).

There is effect of some of the independent variables on the dependent variable (Fatty_type) where showed variable (BMI) very high significant with p_value=0.009, enzyme liver (Got) with probability (p = 0.031) and the degree of a simple effect of sugar control (Hb1Ac) with p_value 0.01, while significant effect of other variables did not show.

The existence of the impact of some independent variables on the dependent variable (Liver_type) where it showed enzyme liver (Got) high impact on increasing the liver size with p_value=0.013, fat cholesterol valued likely would (p=0.013) while the age factor is effect high significant with p_value 0.009, while effect of other variables did not significant.

Robust method: We chosen robust method of analysis to determine the effect of outlier values on the independent variables which assumed some of the values (10%) of them, because the dependent variables are binary data (0,1), the results showed in the following Table 2:

The test value (Likelihood Ratio LR) show an appropriate model used for for analysis (χ²= 71.71 with p <0.0001).

The results in Robust analysis are not different without Robust, for (Fatty_type), where the variable (BMI) has very high effect with
p_value==0.009, enzyme liver (Got) with p_value==0.024, but simple effect of sugar control(Hb1Ac) with p_value=0.068, while no significant effect of other independent variables.

The existence of the impact of some of the independent variables on the dependent variable (Liver_type) as variable enzyme liver (Got) high impact on increasing the liver size with p_value=0.002 and fat cholesterol is significant with p_value=0.016, while the age factor influential degree p_value=0.031, while significant effect of other variables did not show.

Additional statistical tables are needed to show the numbers of the spread of fatty liver and liver size of the study sample according to social standards (Gender, Age group).

Through the above Table 3 it is clear to us that:

### Gender Fatty Type

| Gender | Fattytype | Total |
|--------|-----------|-------|
| 1      | 18        | 44    |
| 2      | 28        | 60    |
| Total  | 46        | 104   | 150

### Age class Liver Type

| Ageclass | Livertype | Total |
|----------|-----------|-------|
| 40-55 years | 23        | 42    |
| 55-70 years | 19        | 43    |
| <40 years     | 8         | 7     |
| >70 years     | 1         | 7     |
| Total         | 51        | 99    | 150

### Age class Fatty Type

| Ageclass | Fattytype | Total |
|----------|-----------|-------|
| 40-55 years | 17        | 48    |
| 55-70 years | 19        | 43    |
| <40 years     | 5         | 10    |
| >70 years     | 5         | 3     |
| Total         | 46        | 104   | 150

**Table 3:** Study sample according to social standards (Gender, Age group).

The spread of fatty liver in the study sample is 69% (104/150) and the proportion of its spread in females 59% (88/150) compared with males, while increasing the liver size in the research sample that 66% have a problem with increasing the size from the normal, and the highest proportions in the age groups (55-70 years; 40-55 years) 43% and 42% respectively.

While the fat spread problems by 46% in the age group (40-55 years) and close to this ratio in the category (55-70 years), and is identical with the medical condition after the injury fat in the previous age group starts the problem of increasing the liver size in next age.

### Conclusions

There is a positive correlation between liver fatty and increase of the size liver responses, but this correlation not significant.

Impact of the explanatory variables (BMI, Age, Hb1Ac, Got, Cho) is approved on liver fatty changes and increased liver size, while do not receive the influence of other explanatory variables on these responses.

Spread of fatty liver disease in females than in males according to the research sample was more prevalent in the age group (40-55 years) while the prevalence of liver size ratio in the age group (55-70 years), a medically acceptable because Category Previous showed the spread of fatty liver disease where after years leads to an increase (inflation) size of the liver.

### Recommendations

Use simulations to test the success of the methods used under several influences (errors distribution type, sample size and assuming a high correlation between the two variables of response values).

More studies are required to study the analysis of logistics multivariable models.

### References

1. Mc Donald BW (1993) Estimating Logistic Regression Parameters for Bivariate Binary Data. Journal of the Royal Statistical Society 55: 391-397.
2. Kimberlee G, Marcello P (1997) The Analysis of Correlated Binary Outcomes Using Multivariate Logistic Regression. Biometrical Journal 39: 309-325.
3. O’Brien SM, Dunson DB (2004). Bayesian Multivariate Logistic Regression. Biometrics 60: 739-746.
4. Yee TW (2008) VGAM Family Functions for Bivariate Binomial Responses.
5. Tabatabai MA, Li H, Eby WM, Kengwoung-Keumo JJ, Manne U, et al. (2014) Robust Logistic and Probit Methods for Binary and Multinomial Regression. J Biomet Biostat 5: 202.
6. Cessie S, Van Houwelingen JC (1994) Logistic regression for correlated binary data. Applied Statistics 43: 95-108.
7. Hosmer D, Lemeshow S (2000) Applied Logistic Regression. John Wiely&sons Ltd, UK.
8. Robert P (2011) Robust Multivariate linear regression using the student t distribution. Tampere University of Technology.
9. Hun MP (2009) Regression Models for Binary Dependent Variables Using Stata, SAS, R, LIMDEP, and SPSS.
10. Stata Press (2007) Stata Base Reference Manual.
11. Rusul FA (2014) Relationship between Type II Diabetes Mellitus and Fatty Liver in Human.
12. Palmgren J (1989) Regression models for bivariate binary responses.