Logit Model of High Blood Pressure

Evelyn N. Okeke, and Joseph U. Okeke

Abstract—This paper is focused on the logit model of high blood pressure. We discussed logit model in detail and applied it to a real life data of high blood and normal blood pressure of 126 individuals. The analysis generated a model for finding the probability of having high blood pressure and revealed that heartbeat rate followed by weight and then age are factors affecting high blood pressure with a concordant percentage of 95.4.

Index Terms—Age; Binary Logit; Essential Hypertension; Heartbeat; Hypertension; Logit Model; Weight.

I. INTRODUCTION

In statistics, logit model or logistic regression is a regression model where the response variable is categorical in nature. When the dependent variable is dichotomous we have binary logistic model but when it has more than two outcome categories we have categorical logistic regression. Binary logit model is appropriate when response takes one of only two possible outcomes representing presence or absence, good or bad, effective or non-effective etc. It is an example of a qualitative response/discrete choice model. Logit model is popularly used in health sciences like epidemiology and market survey analysis to determine the customers’ preference on particular good.

Reference [1] estimated a model of consumer behavior where he examined whether or not an individual had experienced a major negative derogatory report in his/her credit history. This study will develop a model for predicting the probability of presence of high blood pressure in an individual using logit model.

II. HIGH BLOOD PRESSURE

Blood pressure is the force exerted by the blood against the walls of blood vessels, and the magnitude of this force depends on the cardiac output and the resistance of the blood vessels. Blood pressure is determined both by the amount of blood your heart pumps and the amount of resistance to blood flow in your arteries. The more blood one’s heart pumps and the narrower the arteries, the higher the blood pressure. Blood pressure readings are given as two numbers. The systolic pressure (the top number) equals the pressure in the arteries as the heart contracts as its pumps blood around the body. The diastolic pressure (the bottom number) is the pressure as the heart relaxes and refills with blood.

High blood pressure (or hypertension) is a common condition in which the force of the blood against the artery walls is consistently too high that it may eventually cause health problems. Hypertension is defined as having a blood pressure higher than 140 over 90 mmHg (millimeters of mercury), with a consensus across medical guidelines. A diagnosis of hypertension may be made when one or both readings are high. High blood pressure usually has no sign or symptoms, so the only way to know if you have blood pressure is to have your blood pressure measured in the account of [2]. Even without symptoms, damage to the blood vessels, and one’s heart continues and can be detected. High blood pressure is a “silent killer.” When left untreated, the damage that high blood pressure does to one’s circulatory system is a significant contributing factor to heart attack, stroke and other health threats.

About 85 million Americans- one out of every three adults over age 20 have high blood pressure. Nearly one out of six do not even know they have it as stated by [3]. High blood pressure leads to 350 preventable strokes and heart attacks every day in the United Kingdom as observed by [4].

Reference [5] investigated the relationship between age and blood cholesterol on high blood pressure using multiple regression and discovered that age and blood cholesterol have significant relationship with systolic blood pressure. Reference [6] examined the distribution of systolic blood pressure by means of maximum likelihood fitting of various statistical models and found evidence for a compound or mixed log-normal distribution.

High Blood Pressure Causes and Effects

According to [7] high blood pressure is dangerous because it makes the heart work harder to pump blood out to the body and contributes to hardening of the arteries or atherosclerosis, to stroke, kidney disease, and to the development of heart failure. They further stated that weight, diet and age are causes of high blood pressure.

Reference [8] and [3] gave the ranges for learning about and checking one’s blood pressure numbers and what they mean as:

<table>
<thead>
<tr>
<th>TABLE I: BLOOD PRESSURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood Pressure Category</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Normal</td>
</tr>
<tr>
<td>Pre-hypertension</td>
</tr>
<tr>
<td>High blood pressure</td>
</tr>
<tr>
<td>(hypertension) stage 1</td>
</tr>
<tr>
<td>High blood pressure</td>
</tr>
<tr>
<td>(hypertension) stage 2</td>
</tr>
<tr>
<td>Hypertensive Crisis</td>
</tr>
<tr>
<td>(Emergency care needed)</td>
</tr>
</tbody>
</table>

According to [9], high blood pressure whose underlying cause cannot be determined is called “essential hypertension.” They claimed that though this type of hypertension remains somewhat mysterious but it has been linked to certain risks factors. They further stated that high
blood pressure tends to run in families and is more likely to affect men than women. Also they claimed that age and race play a role in determining hypertension. They emphasized that blacks are twice as likely as white to have high blood pressure in the United States, although the gap begin to narrow around age 44. Finally, they stated that at age 65, black women have the highest incidence of high blood pressure.

Reference [9] also stated that several factors and conditions may play a role in high blood pressure development. The factors according to them includes: smoking, being overweight or obese, lack of physical activity, too much salt in the diet, too much alcohol consumption, stress, older age, genetics, family history of high blood pressure, chronic kidney disease, adrenal and thyroid disorder, sleep apnea, exposure to ozone, pregnancy, ethnic background, and some aspect of gender.

III. LOGIT MODEL

Logit and probit models are among the most popular models of binary response variable. In logit model the dependent variable is a binary response (like yes or no, good or bad, defective or non-defective, present or absent, responsive or non-responsive response etc), commonly coded as a 0 or 1 variable. When the dependent variable Y is dichotomous, the only values it can have are 1 and 0, that is, \( y_i \in \{0,1\} \) with probabilities \( p_i \) and \( 1 - p_i \) respectively. In this study, \( y_i = 1 \) is a situation where an individual is having high blood pressure and \( y_i = 0 \) if he has normal blood pressure.

Suppose we have a model

\[
Y = X\beta + \epsilon
\]

(10)

the linear regression predictor will look like

\[
\hat{Y} = X\beta
\]

and a link function is a function linking the actual \( Y \) to the estimated \( \hat{Y} \) in a statistical model. In a more general form, a link function is some function \( F(\cdot) \), such that,

\[
F(Y) = \hat{Y} = X\beta
\]

(2)

For logit model the link function

\[
F(Y) = \log(Y)
\]

(3)

Both logit and probit predictors can be written as

\[
\hat{Y} = f(X\beta)
\]

One problem with (1) which is often estimated using OLS is that the \( Y \) on the left-hand-side has to either zero or one but the linear predictor \( X\beta \) on the right-hand-side can take real value, so there is no guarantee that the predicted values will be in the correct range (that is, the regression line may lead to prediction outside the range of zero and one) unless complex restrictions are impose on the coefficient. The functional form of the regression line by OLS will assume that \( ith \) and \( jth \) observations have the same marginal effect, which is probably not appropriate for binary response. Again a residuals plot would quickly reveal heteroscedasticity in the account of [10]. A simple solution to these problems is to transform the probability to remove the range restrictions, and model the transformation as a linear function of the covariate. Therefore we need to transform the dichotomous \( y_i \in \{0,1\} \) into a continuous, real-valued \( y_i^* \in (-\infty, \infty) \). So we need a link function \( F(Y) \) that takes a dichotomous variable \( Y \) and give us a continuous variable \( Y' \) so that we can run:

\[
F(Y) = Y' = X\beta + \epsilon
\]

(4)

The random variable \( Y \) can be transformed to real line using standard normal distribution but apart from this, odds ratio can also be used to transform the dichotomous response variable \( Y \). If the realization of a random variable occurs with probability \( p \), then the odds of it happening is

\[
O(p) = \frac{p}{1-p}
\]

(5)

The odds of the \( p \) can transform \( p \) from the interval \([0, 1]\) to the half-line \([0, \infty)\). The odd ratio is always non-negative. The logarithm of the odds ratio \( \log(O(p)) \) can transform \( p \) to the entire real line \((-\infty, \infty)\) as \( p \) varies between 0 and 1 which is completely continuous. The log of the odds ratio of \( p \) is called logit function of \( p \), that is,

\[
F(p) = \log(O(p)) = \log\left(\frac{p}{1-p}\right)
\]

(6)

The density function associate with it is very close to a standard normal distribution. If we take the \( \log(O(p)) \) as a link function, then

\[
\log\left(\frac{p}{1-p}\right) = X\beta
\]

(7)

as in two above. Therefore, we can say that logit function is link function because it “links” the probability to the linear function of the predictor variable.

Equation (7) can translates back to the original \( Y \) as

\[
\log\left(\frac{p}{1-p}\right) = X\beta
\]

(8)

Taking the log inverse of both sides we have that

\[
\frac{p}{1-p} = e^{X\beta}
\]

\[
\frac{p}{1-p} = e^{X\beta}(1-p)
\]

\[
p = e^{X\beta} - pe^{X\beta}
\]

\[
p(1 + e^{X\beta}) = e^{X\beta}
\]

\[
p(y = 1|x) = \frac{e^{X\beta}}{1 + e^{X\beta}}
\]

(9)

The resultant equation (8) is the chance or probability of having high blood pressure.

Reference [11] showed that the estimator \( \beta \) could be calculated with the formula:

\[
\hat{\beta} = \arg \max_{\beta} \left\{ \sum_{i=1}^{n} \left[ y \ln\left(\frac{e^{X\beta}}{1+e^{X\beta}}\right) + (1 - y) \ln\left(\frac{1}{1+e^{X\beta}}\right)\right]\right\}
\]

(9)
Similarity and Dissimilarity of Logit and Probit Model

Logit and probit differ in how they define the link function $F(\cdot)$. The logit model uses something called the cumulative distribution function of the logistic distribution. The probit model uses something called the cumulative distribution of the standard normal distribution to define $F(\cdot)$. Both functions will take any number and rescale it to fall between 0 and 1. Hence, whatever $XB$ equals, it can be transformed by the function to yield a predicted probability. Any function that would return a value between zero and one would do the trick, but logistic and standard normal cumulative density functions turn out to be convenient mathematically and are programmed into most of the common statistical programs as claimed by [12] and [10].

IV. DATA AND ITS INTERPRETATION

The data used in this study is real life data collected from the hospital records of Nnamdi Azikiwe Teaching Hospital Nnewi. Information on age, height, weight and heartbeat rate/pulse of 65 people with high blood pressure and 61 persons with normal blood pressure were collected. From the data we observed that people within the age of 21-37 that have high blood pressure are those that have high weight. People from age 55 and above that have normal blood pressure are those whose weights are on the average.

The principal component analysis of the four factors collected was computed as a preliminary analysis to check for possible correlate and also to know the percentage of total variation each factor explained in the original data. We observed from PC analysis that age and weight are major factors that affect high blood pressure.

V. RESULTS AND THEIR INTERPRETATION

The logit regression analysis of the data revealed that heartbeat rate followed by weight and then age are factors affecting high blood pressure with a concordant percentage of 95.4. The resultant probability (which is the logit model) of high blood pressure from the analysis is

\[
P = \frac{e^{-43.11 + 0.03494 \text{Age} + 0.1432 \text{Weight} + 0.1031 \text{Height} + 0.2035 \text{Heartbeat rate}}}{1 + e^{-43.11 + 0.03494 \text{Age} + 0.1432 \text{Weight} + 0.1031 \text{Height} + 0.2035 \text{Heartbeat rate}}}
\]

The goodness-of-fit test of the model by Pearson has a p-value of 0.000 which together with concordant percentage and deviance p-value of 1.000 indicate that the resulting model is significant in predicting the chance of one having high blood pressure.

From the model we calculated the odds ratio of every individual observation in the data. The calculated odds ratio of individuals showed that people with very high weight have high odds value indicating high chance of having high blood pressure.

ACKNOWLEDGMENT

We extend our sincere appreciation to all the authors whose works are referenced in this paper for their findings and contributions in the subject matter that charted the way for further research.

REFERENCES


She was a LECTURER at the Nnamdi Azikiwe University (NAU, Awka) 2001-2013. Presently, she is the Head of the Department of Mathematics and Statistics of the Federal University Wukari, Taraba State, Nigeria. (e-mail: evelyn70ng@yahoo.com). She has published in both local and foreign reputable journals. Her research interests is in the area of Discriminant Analysis.

Dr.Mrs. Okeke is a member of the Nigerian Statistical Association, a consultant with the United Nations Development Program, the Secretary, Anambra State University Wukari, Nigeria. Chairperson Welfare committee, Faculty of Pure and Applied Sciences, Federal University Wukari, Nigeria.


He was a LECTURER at the Anambra State University now Chukwuemeka Odumegwu Ojukwu University (2007-2013). Presently, he lectures in the Department of Mathematics and Statistics of the Federal University Wukari, Taraba State, Nigeria (e-mail:uche70ng@yahoo.com). He has published in both local and foreign reputable journals. His research interests are in the areas of Econometric dynamic modeling and Multivariate classification modeling.

Dr. Okeke is a member of the Nigerian Statistical Association, a consultant with the United Nations Development Program, the Secretary, Anambra West Elite Club (2012 to date), the seminar coordinator(2015-2017) Faculty of Pure and Applied Sciences, Federal University Wukari, Nigeria.

DOI: http://dx.doi.org/10.24018/ejers.2017.2.12.429