International Journal of Modelling & Applied Science Research
Published by Cambridge Research and Publications

Vol. 28 No.9
June, 2023.

LOGISTIC REGRESSION ANALYSIS TO PREDICT THE DISCHARGE STATUS OF HYPERTENSIVE PATIENTS IN HOSPITAL

FASASI, S.K. & OSUNTOKI, N.B.

Department of Mathematics and Statistics, School of Science, the Oke – Ogun Polytechnic, Saki, Oyo State

Abstract

The model summary gives us the usefulness of the model and provides the -2 log likelihood and pseudo R^2 values for the full model. The Cox & Snell R Square and Nagelkerke R Square values shows the amount of variation in the dependent variable explained by the model (explanatory variables) from 0 – 1. In the table above 5.5% of the variation in the dependent variable is explained by the logistic model. Nagelkerke R Square is 10% relationship between the predictors and prediction. -2log likelihood is 641.414 which is high and indicates a poor prediction of the outcome because the smaller the statistic the better the model. -2LL value for this model (641.414) is what was compared to -2LL for the previous null model in the omnibus test of model coefficient which told us that there was significant decrease in -2LL, that is, our new model (with explanatory variables) is significantly better fit than the null model.

Keywords: log likelihood, null model, logistic model, Nagelkerke R Square

INTRODUCTION

Hypertension or High Blood Pressure (BP) sometimes called Arterial hypertension is a chronic medical condition in which the blood pressure in arteries rises. High BP can also be defined as the force of blood pushing against blood vessel walls. It is a condition that occurs as a result of high blood pressure above 140 over 90mmHg. High blood pressure is a major risk factor for heart disease and stroke. It is one of the several factors associated with cardio vascular disease, BP lowers during sleep and rises in awakening in response. Hypertension can occur in children or adults, but it is more common among people over 40years of age. Hypertension is widely spread in Africa, among the middle age, elderly ones, overweight, alcoholics, women who take oral contraceptives. People with kidney disease, diabetes are likely to suffer high blood pressure which eventually generates hypertension. Blood pressure is summarized by two measurements, systolic and diastolic which depends on whether the heart muscle is contracting (systole) or relaxed (diastole). Normal BP is within the range of 100-140mmHg systolic and 60-90mmHg diastolic. High blood pressure is present if it is often or above 140/90mmHg. By medical advices, the normal blood pressure for most adults is less than 120/80mmHg. Hypertension cannot be cured but can be managed, it does not cause symptoms but in the long terms leads to complicated cases by narrowing of blood vessels. Doctors predict high blood pressure over a number of visits using a sphygmomanometer which is applying an inflatable cuff to the upper arm. Drug treatment can be used if ones blood pressure is at or above 140/90mmHg. One of the most common causes of treatment resistant hypertension is primary aldosteronism. There are various risk factors that contribute to the development of hypertension. These risk factors either individually results in the development of hypertension or when its associates with other risk factors. Primary hypertension has no identifiable cause but it is has several risk factors such as age, high salt intake, low potassium diet, sedentary lifestyle, stress, genetic variation, obesity , hormonal changes, alcohol, renin elevation, vitamin deficiency. Secondary hypertension is a high blood pressure that has an underlying cause, this includes: kidney diseases, renal disorders, tumors, endocrine disorders, adrenal cortical abnormality, hormonal contraceptives, sleep apnea, pregnancy, cancers, drugs, etc. people with diabetes, psoriasis, pregnancy, high cholesterol, post-menopausal women have a higher risk of having high blood pressure.

Material and Methodology

The sole aim of this study is to examine the factors contributing to the discharge status or outcome of hypertensive patients using binary logistic regression analysis approach and also to determine association and strength among independent variables or parameters, this includes sex, age, occupation and length of stay in a hospital.

This study is designed to present an extension of applying logistic regression technique in predicting the outcome of hypertensive patient (dead or alive). Logistic regression is used to predict a categorical (usually dichotomous) variable from a set of predictor variables. Logistic regression is used to describe the data and explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. Logistic regression is not limited to medical research; it is valuable in business world (e.g. Consumers decisions to buy or not buy), education (e.g. pass or fail a subject) etc. and in this particular study, a hypertensive patient may either live or die. In summary, this study illustrates the scope/method of logistic regression and how it can be applied to real life situations by selecting a case study in Muslim hospital, Saki, Oyo state, Nigeria.

This research work is limited to the reported cases in Muslim Hospital, Saki, Oyo state. The variables considered are age, survival, length of stay in hospital, occupation and sex.

The appropriate hypotheses are

To examine the hypothesis that all the regression coefficients for the independent variables are zero. That is;

H0 : B1=B2=B3=B4=0

H1 : Not H0 (at least one of the coefficient of the independent variable is not zero)

Where B1, B2, B3, B4 are the coefficients of the independent variables, age, sex, occupation, and length of stay respectively.

The Logistic Equation

Logistic regression equation is given as

$$\text{Logit(p)} = \log\left(\frac{p}{1-p}\right) =$$

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u_k$$

To solve this equation for p, we first apply the exponential function to both sides of the equation:

$$\exp\left(\log\left(\frac{p}{1-p}\right)\right) = \exp\left(\alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\right).$$

Recall that exp(z) $= e^z$ so that the right hand side of the above equation is

$$\exp\left(\alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\right) = e^{(\alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}$$

Also remember that "log" is the natural logarithm, so the exponential function is its inverse. Thus, the left hand side is

$\exp\left(\log\left(\frac{p}{1-p}\right)\right) = \left(\frac{p}{1-p}\right)$

Thus, after exponentiating both sides, logistic regression equation becomes:

$$\frac{p}{1-p} = \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k).$$

$$= e^{(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)}$$

Next multiply both sides by $1 - p$

$$p = (1-p)e^{(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)}$$

and then "break up" the (1-p) term,

$$p = e^{(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)} - pe^{(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)}$$

Now move the $pe^{(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)}$ term (the last term on the right-hand side) over to the left-hand side by adding it to both sides:

$$p + pe^{(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)} = e^{(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)}$$

Next, factor out the p,

$$p\left(1 + e^{\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k}\right) = e^{(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)}$$

Finally, divide both sides by $(1 + e^{\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k})$ to get p:

$$p = \frac{e^{\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k}}{1 + e^{\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k}}$$

There is one other form of this equation that is commonly used. It is obtained by multiplying the top and bottom of the right hand side of the equation for p by $e^{-(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)}$. Since $e^0e^0 = 1$, this gives

$$p = \frac{e^{\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k}}{1 + e^{\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k}} \times \frac{e^{-(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)}}{e^{-(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)}}$$

$$= \frac{e^{\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k} \times e^{-(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)}}{e^{-(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)} + e^{-(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)} \times e^{\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k}}$$

$$= \frac{1}{e^{-(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)} + 1}$$

The terms in the denominator are customarily written in the opposite order. So the second form of the equation for p is

$$p = \frac{1}{e^{-(\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k)} + 1}$$

P = the probability that a case is in a particular category

α = constant of the equation

β = coefficient of the predictor variables

Thus, the logistic regression involves fitting an equation of the form to the data;

Logit (p) = $\log(\frac{p}{1-p}) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u_k$

Results and Discussion

Variables in the Equation

This table tells us about the importance and contribution of each of the explanatory variables. It shows the logistic coefficient (β) , the closer a logistic coefficient is to zero, the less influence in the prediction of the logit, the Wald statistic gives an index of the significance of each predictor in the equation, if p-values is less than 0.05, reject the null hypothesis and conclude that the variable make a significant contribution. Age (0.000), sex (0.008), length of stay (lsh) (0.001) are less than 0.05 that is, they contributed significantly to the prediction while occupation (0.581) does not make any significant contribution to the prediction. The explanatory variable occupation is dropped since the effect is not significant.

The EXP (B) shows the extent to which raising the corresponding measure by one unit influences the odd ratio. If value exceeds 1 then the odds of an outcome increases otherwise decreases.

The odds ratio for Age is 0.966 which is less than 1, that is, any unit increase in the age of the patient, there would be a drop of being discharged alive (95% CI: 0.952 to 0.980, p-value of $0.000 < 0.05$) when sex, occupation and lsh are held constant, the β value of -0.35 means the patient are of younger age and likely to be discharged alive holding other independent variables fixed.

The odds ratio for sex is 0.553 which is less than 1, that is, any unit increase in the predictor (female) there would be a decrease in the odds of being discharged alive (95% CI: 0.356 to 0.859, p value of $0.008 < 0.05$) the β value of -0.592 means the male has the higher odds of being discharged alive than the female holding other independent variables fixed.

The odds ratio for occupation is 1.138 which is greater than 1 which implies that for any unit increase in the predictor (unemployed) there would be an increase in the odds of being discharged alive at (95% CI: 0.719 to 1.851, p value of $0.581 < 0.05$) the β value of 0.129 implies that unemployed patients have an higher probability of being discharged alive holding other predictors fixed.

The odds ratio for length of stay is 2.310 which is greater than 1 implies for any unit in this predictor, there would be an increase in the odds of being discharged alive at (95% CI: 1.438 to 3.713, p value of $0.001 < 0.05$) . The β value of 0.837 implies that the longer days a patient spend in a hospital the higher the probability of been discharged alive.

Thus, the logistic regression model is:

$$\text{Log}\left(\frac{p}{1-p}\right) = 2.330 - 0.35(age) - 0.592(sex) + 0.129(occupation) + 0.837(lsh)$$

Table 1 Variables in the Equation

	B	S.E	Wald	df	Sig.	EXP(B)	95% C.I for EXP Lower	Upper
Step 1[a] age	-0.35	0.008	20.775	1	0.000	0.966	0.952	0.980
sex	-	0.225	6.944	1	0.008	0.553	0.356	0.859
	0.592	0.234	0.305	1	0.581	1.138	0.719	1.801
occupation	0.129	0.242	11.976	1	0.001	2.310	1.438	3.713
lsh	0.837	0.610	14.563	1	0.000	10.275		
	2.330							
constant								

a. Variables entered on step 1, age, sex, occupation, lsh.

Model Summary

The model summary gives us the usefulness of the model and provides the -2 log likelihood and pseudo R^2 values for the full model. The Cox & Snell R Square and Nagelkerke R Square values shows the amount of variation in the dependent variable explained by the model (explanatory variables) from 0 – 1. In the table above 5.5%

of the variation in the dependent variable is explained by the logistic model. Nagelkerke R Square is 10% relationship between the predictors and prediction. -2log likelihood is 641.414 which is high and indicates a poor prediction of the outcome because the smaller the statistic the better the model. -2LL value for this model (641.414) is what was compared to -2LL for the previous null model in the omnibus test of model coefficient which told us that there was significant decrease in -2LL, that is, our new model (with explanatory variables) is significantly better fit than the null model.

Table 2 Model Summary

Step	-2 log likelihood	Cox & Snell R square	Nagelkerke R square
1	641.414[a]	0.055	0.100

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001

Conclusion

Logistic regression importance in analysis cannot be overemphasized and because of its extensive use in medical research it has been employed in this study. This study is concluded that the occupation is not necessarily important but the older the patient the slimmer the chances of being discharged alive and the more days spent in the hospital, the more the chance of survival and male has the more likelihood of being discharged alive.

Recommendation

In the course of this study, age, sex and length of stay affects the discharged outcome of hypertensive patients. These factors affecting the discharged status should be looked into. We recommend that awareness should be created through social media and any other available means on the need for regular medical checkup and the precautions to take in other not to fall a victim.

References

Adediran O.S, O. I. (2013). Hypertension Prevalence in an Urban and Rural Area of Nigeria. *Journal of Medicine and Medical Sciences*, vol. 4(4) pg. 149-154.

Adeloye D, Basquill., & Aderemi A.V, Thompson J, Felix A. (2014). An Estimate of the Prevalence of Hypertension in Nigeria. *Article in Journal of Hypertension*.

Alexander, A. K. *Logistic Regression Analysis of Factors Associated with Hypertension Prevalence*. Institute of Distance Learning Kwame Nkumah University of Science and Technology, Ghana.

Chao-Ying J. P, K. L. (2002). An Introduction to Logistic Regression and Reporting. *Journal of Educational Research*, vol.96 (No.1).

Ekwunife O.I, A. C. (2011). A Meta Analysis of Prevalence rate of Hypertension in Nigeria Population. *Journal of Public Health and Epidemology*, vol. 3(13) pp 604-607.

Ezekwesili C.N, O. C. (2016). Epidemiological Survey of Hypertension in Anambra State, Nigeria. vol 19 issues 5 pg. 659-667.

Fall. (2015). *Newsom Data Analysis II; More on Model Fit and Significance of Predictors with Logistic regression.* Portland.

Hajjar I, & Kotchen J.M, T. A. (2006). Hypertension: Trends in Prevalence, Incidence and Control. *Annual Review of Public Health*, vol.27 pg. 465-490.

Kate, O. O. *Knowledge of Hypertension among Adults in Owerri Senatorial Zone of Imo State, Nigeria.* University of Nigeria, Nsukka, Nsukka.

Modele O. Ashaye, W. H. (2003). Hyprtension in Blacks: A Literature Review . *Ethnicity and Disease*, vol.13.

National Heart foundation of Australia. (2016). *Guideline for the diagnosis and Management of Hypertension in Adults.* Melbourne.

Phil Reed, Y. W. (2013). Logistic Regression for Risk Factor Modelling in Stuttering Research. *Journal of Fluency Disorders*, vol. 38, 88-101.

Walters T.A, C. A. (2016). Hypertension, an Emerging Problem in Rural Cameroon: Prevalence, Risk factors and Control . *International Journal of Hypertension*.

Wan Muhamad Amir W.A, M. A., & Mohammed P.H, Z. A. (2014). Association of Hypertension with Risk Factors Using Logistic Regression. *Applied Mathematics Science*, vol.8 2563-2572.