

Assessing Discriminatory Performance of a Binary Logistic Regression Model

Noora Shrestha

Department of Mathematics and Statistics

Tribhuvan University

Kathmandu, Nepal

ABSTRACT

The evaluation of fitted binary logistic regression model is very important in assessing the appropriateness of a model for specific purposes. The study proposes to assess the discriminatory performance of a binary logistic regression model to correctly classify between the cases and non-cases. The discriminatory performance of binary logistic regression model is measured using two approaches. The first approach is the use of fitted binary logistic regression model to correctly predict the subjects that are cases and non-cases, with the help of the parameters sensitivity and specificity. The alternative approach is based on receiver operating characteristic (ROC) curve for the fitted binary logistic regression model and then determining the area under the curve (AUC) as a measure of discriminatory performance. The value of sensitivity is observed to be greater than the value of 1- specificity, which signifies suitable discrimination for the mentioned cut point. The area under the curve indicates that there is evidence of reasonable discrimination reported by the fitted model.

Key Words: Binary Logistic Regression, Discriminatory Performance, Sensitivity, Specificity, Receiver Operating Characteristic (ROC), Area Under the Curve (AOC).

1. INTRODUCTION

Making predictions have developed in every part of life with the application of statistical models to predict the expected occurrence of the event for example weather forecasts, health outcomes, and forecasts in business organization. The prediction helps people and organizations to decide their future plans and therefore, assessing accuracy of the forecasts is very important for decision-making [1, 2]. The application of binary logistic regression model to predict the classification of subjects as cases and non-cases are extremely significant in many researches such as distribution of species [3], health outcome [4]. The binary logistic regression model is suitable for a dichotomous categorical response variable and the predictors can be either quantitative or categorical. The evaluation of the fitted logistic model using independent data is very important in assessing the appropriateness of a model for specific purpose [5].

The receiver operating characteristic (ROC) curve is a statistical tool to measure the correctness of likelihoods of an event. ROC curve provides a comprehensive way to analyze the accuracy of predictions. The procedures for assessment of accuracy depend on the type of the predictor. The area under the ROC curve has a significant clarification for disease categorization from healthy subjects [6]. The sensitivity and specificity of a test is vital to assess its application for identification of case and non-cases. The sensitivity of a diagnostic test is the proportion of subjects for whom the result is a case that are correctly classified. The specificity is the proportion of subjects for whom the result is non-case that are correctly classified by the test [7].

The objective of the proposed study is to assess the discriminatory performance of a binary logistic regression model to correctly classify between the cases and non-cases. The discriminatory performance of binary logistic regression model is measured using two approaches. The first approach for measuring the discriminatory performance is the use of fitted binary logistic regression model to correctly predict the subjects that are cases ($y=1$) and non-cases ($y=0$) with the help of the parameters sensitivity and specificity. The alternative approach is based on the Receiver operating characteristic (ROC) curve for the fitted logistic regression model and then determining the area under the curve as a measure of discriminatory performance [8].

2. MATERIALS AND METHODS

The binary logistic model has to be fitted for the dataset containing information about two hundred and forty five female subjects of age group between eighteen to twenty four years and used to discriminate the observed overweight subjects ($y=1$) from the observed not overweight subjects ($y=0$). The approval of the study was received from the ethical and research committee of the educational institute and the written consent of the subjects were taken prior. The researcher on the basis of the guidelines of the World Health Organization noted the physical measurements of the subjects.

According to the report of World Health Organization, the subjects with neck circumference, $NC \geq 34$ cm were considered as overweight and waist to height ratio, $WHTR \geq 0.49$ were grouped as overweight and whereas otherwise, not overweight. In this study, the binary outcome variable overweight ($y=1$) or not overweight ($y=0$) was classified on the basis of body mass index (BMI) [9, 10]. The subjects with $BMI \geq 25$ kg/m² were grouped as overweight and subjects with $BMI < 25$ kg/m² grouped as not overweight. The purpose of using a logistic regression model was to assess whether the predictors neck circumference and waist to height ratio (categorical) could be used and significant as a screening measure to determine the case of overweight [10,11].

2.1 Binary Logistic Regression Model

For k independent variables and $i= 1,2,3,\dots, n$ independent observations, the logistic model is given by

$$\text{Log Odds or Log} \left[\frac{p_i}{1 - p_i} \right] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

Where, p_i is the probability that $y=1$ and x_{ik} are categorical and continuous explanatory variables. The log odds equation can be solved for the prediction of probability as

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})}$$

Where, \exp is the exponential function (e^x) [12, 13].

2.1.1 Measuring Discrimination Performance

The discriminatory performance of fitted binary logistic regression model is assessed, by studying the agreement between predicted outcome and the observed outcome, with the help of 2x2 classification table. A subject is predicted to be a case or non-case based on whether the predicted probability is greater or less than a specified threshold probability value. The 2x2 general classification table is given below:

Table1. General classification table

Predicted Outcome	Observed Outcome			
		Y=1 (Case)	Y=0 (Non-case)	
Y=1 (Case)		TP (A)	FP (B)	A + B
Y=0 (Non-case)		FN (C)	TN = (D)	C + D
		$n_1 = A + C$	$n_0 = B + D$	A+B+C+D

Where, TP is true positives, FP is false positives, TN is true negatives, and FN is false negative [3, 8, 14]. The table can be used to calculate the two major parameters sensitivity and specificity and other indices and they are defined as:

$$\text{Sensitivity (Se or TPF)} = \frac{\text{The number of cases correctly predicted (TP)}}{\text{Total number of cases in the sample (n}_1\text{)}} = \frac{A}{A + C}$$

$$\text{Specificity (Sp or TNF)} = \frac{\text{The number of non - cases correctly predicted (TN)}}{\text{Total number of non - cases in the sample (n}_0\text{)}} = \frac{D}{B + D}$$

$$\text{False Positive Fraction(FPF)} = \frac{\text{The number of false cases predictions (FP)}}{\text{Total number of non - cases(n}_0\text{)}} = \frac{B}{B + D}$$

$$\text{False Negative Fraction(FNF)} = \frac{\text{The number of false non - cases prediction (FN)}}{\text{Total number of cases in the sample (n}_1\text{)}} = \frac{C}{A + C}$$

The accuracy of the fitted model is calculated as

$$Accuracy = \frac{A + D}{A + B + C + D}$$

2.1.2 Receiver operating characteristic (ROC) curve

Receiver operating characteristic curve, a probability curve, is a performance measurement for categorization issues at different threshold settings. ROC curve is used to determine the suitable cutoff values for categorization by comparing the sensitivity and specificity of various cutoff values. The ROC curve is the entire set of possible true positive fraction (TPF) and false positive fraction (FPF) attainable by dichotomizing the outcome variable y with different thresholds. The ROC curve is $ROC(c) = \{(FPF(c), TPF(c)), c \in (-\infty, \infty)\}$. When the threshold c increases, both FPF(c) and TPF(c) decrease [12,13,15]

ROC curve is used to predict a binary outcome that tells how well the fitted model separates the true cases from the true non-cases. This prediction will be done getting the area under the curve (AUC) which is a measure of discriminatory performance of a fitted model. The area under the ROC curve is $AUC = \int_0^1 ROC(t)dt = P[(y = 1) > P(y = 0)]$. If the AUC value is close to one or $Se \geq 1 - Sp$, the model is excellent to classify between the cases and non-cases. If AUC is 0.5 or $Se = 1 - Sp$, the model is worthless and not capable to distinguish between the cases and non-cases. There will be positive discrimination if AUC lies between 0.5 and 1.0 and negative discrimination if AUC lies between 0 and 0.5. Hence, the fitted model is expected to provide a proper discrimination if the true cases have a greater predicted probability than the true non-cases [15,16,17].

3. RESULTS AND DISCUSSION

The binary logistic regression model was fitted for the dataset satisfying the assumption of categorical outcome variable, the independent variables were categorical and continuous, and the independence of observations using SPSS. The output of the model depicts 48 subjects were classified in overweight group and 197 subjects were classified as not overweight with 80.4% accuracy in the model when the explanatory variables were not included. The first approach for assessing the discriminatory performance is the use of fitted binary logistic model to correctly predict the subjects that are overweight (y=1) and not overweight (y=0) and then resolve the proportions of parameters sensitivity and specificity.

The table 2 reports the output of logistic model with significant predictors. The model delivers good discriminatory performance, if the covariates neck circumference (NC) and waist to height ratio categorical (WH_cat) included in the model.

Table2. Fitting logistic regression model

Parameter	Estimate (B)	Std. Err.	Wald Chi Sq.	df	Sig.	Exp(B)	95% C.I. for Exp (B)	
							Lower	Upper
NC	0.323	0.111	8.45	1	0.004	1.382	1.111	1.719
WH_cat	-1.744	0.402	18.79	1	0.000	0.175	0.079	0.384
Constant	-11.072	3.785	8.56	1	0.003	0.000		

The fitted logistic regression model can be expressed as

$$\text{Log} \left[\frac{p_i}{1 - p_i} \right] = \beta_0 + \beta_1 * NC + \beta_2 * WH_cat = -11.072 + 0.323 * NC - 1.744 * WH_cat$$

The fitted model helps to predict which subjects will be overweight and which will not be overweight. If the predicted probability of subjects is greater than or equal to 0.5, it can be predicted that the subject will be overweight and else, not overweight. The classification table 3 shows the observed and predicted outcomes of the subjects at a cut point Cp= 0.5.

Table3. Classification Table

Predicted Outcome	Observed Outcome	
	Y=1(Overweight)	Y=0 (Not Overweight)
	Y=1(Overweight)	TP = 14
Y=0 (Not Overweight)	FN = 34	TN = 190
	n ₁ = 48	n ₀ = 197

Here, the numbers of true or observed overweight subjects that are predicted to be overweight are true positives (TP = 14) and the number of observed not overweight subjects that are predicted to be not overweight are true negatives (TN = 190) with 83.3% if the accuracy of the model. The sensitivity is the proportion of true positives among all the subjects that are

overweight and specificity is the proportions of true negatives among all the not overweight subjects. If both sensitivity and specificity values are equal to 1, then the condition of perfect discrimination will be obtained. The sensitivity (Se) = TP/ n₁ = 14/48 = 0.292 or 29.2% and specificity (Sp) = TN/ n₀ = 190/197 = 0.964 or 96.4%. The false positive count is FP = 7. The expected value of 1- Specificity = FP/ n₀ = 7/ 197= 0.036 which is close to zero, and hence Se= 0.292 is expected to be greater than the value of 1-Sp, which signifies the suitable discrimination for this cut point.

The problem of measuring discrimination is that the two values, sensitivity and specificity may vary on the basis of cut-points selected which is the drawback to measuring discrimination using the first approach. Hence, an alternative approach to assess the discriminatory performance is applied.

The alternative approach is based on receiver operating characteristic (ROC) curve for the fitted binary logistic model and then determining the area under the curve as a measure of discriminatory performance. The ROC curve for two different predictors is shown in the figure1. The ROC curve in the figure, when applied to a logistic model, is a graph of sensitivity and 1-specificity obtained from the range of cut points for the predicted value. The value of 1-specificity gives the proportion of observed not overweight subjects that are falsely predicted to be overweight. The highest value of sum of sensitivity and specificity gives optimum cutoff value. The ROC curve begins at the origin (0,0), y-axis takes the value from 0 to 1, and x-axis takes the value from 0 to 1. The diagonal line joining (0,0) and (1,1) is the reference line. The figure 1 shows that the neck circumference does not deliver a suitable sign of overweight than the waist to height ratio.

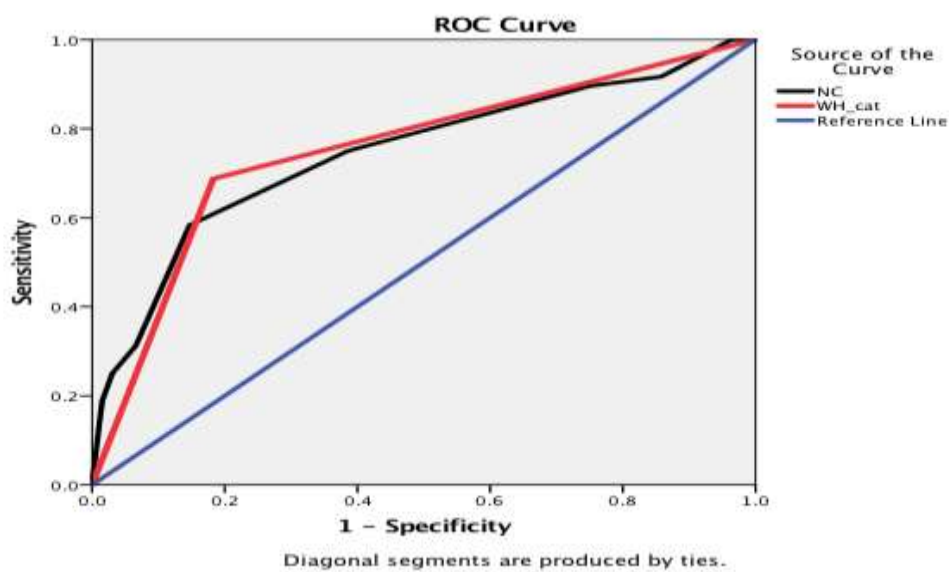


Figure 1. Receiver operating characteristic (ROC) curve

If the area under the curve (AUC) is larger, then the model discriminates better and hence the area under the curve is calculated and presented in the table 4.

Table4. Area under the curve

Variables	AUC	Std. Err.	p-value	95% Confidence Interval	
				Lower	Upper
WH_cat	0.752	0.042	0.000	0.670	0.835
NC	0.745	0.044	0.000	0.659	0.832

The area under the ROC curve for WH_cat and NC are 0.752 and 0.745 respectively. It means there is evidence of 75.2% and 74.5% chance that the logistic model is capable to distinguish between overweight and not overweight subjects. The AUC for WH_cat indicates the probability that a female subject will be overweight has a waist to height ratio greater than that of a subject who is not overweight. The p-value is less than 0.0001 and the 95% confidence interval for AUC is 0.670 to 0.835 for WH_cat predictor, suggesting that waist to height ratio may deliver a better predictive test for overweight of a subject. The significant AUC value of neck circumference is very close to AUC of WH_cat, so it also delivers a good predictive test for overweight of a subject.

4. CONCLUSION

The receiver operating characteristic curve suggests a suitable process to assess the discriminatory performance of a fitted logistic model. The value of sensitivity is observed to be greater than the value of 1- Specificity, which signifies appropriate discrimination for the mentioned cut point. In practice, it is usually difficult to obtain the area under the curve ≥ 0.9 . The observed area under the curve for the variables waist to height ratio and neck circumference indicates that there is evidence of reasonable discrimination subjects to overweight and not overweight group using the fitted logistic model. However, there are other modern techniques and methods that signifies higher accuracy and widely applicable in the research.

REFERENCES

- [1] J.E. Thornes, and D.B. Stephenson, "How to judge the quality and value of weather forecast products," *Meteorological Applications*, vol.8 (3), pp. 307–314, September 2001.
- [2] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, New York, 2003.
- [3] J. Pearce, and S. Ferrier, "Evaluating the predictive performance of habitat models developed using logistic regression," *Ecological Modelling*, vol.133, pp. 225-245, 2000, 10.1016/S0304-3800(00)00322-7.
- [4] S. Noora, "Application of binary logistic regression model to assess the likelihood of overweight," *American Journal of Theoretical and Applied Statistics*, vol. 8, No. 1, pp. 18-25, 2019, doi: 10.11648/j.ajtas.20190801.13
- [5] D.P. Allison, *Logistic Regression Using SAS[®]: Theory and Application, Second Edition*, SAS Institute Inc., Cary, NC, 2012.
- [6] K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Caspian Journal of Internal Medicine*, vol. 4(2), pp. 627-35, 2013.
- [7] V. Bewick, L. Cheek, and J. Ball, "Statistics review 13: receiver operating characteristic curves," *Critical care (London, England)*, vol. 8 (6), pp. 508-12, 2004, doi:10.1186/cc3000.
- [8] D.G. Kleinbaum, and M. Klein, *Logistic regression, statistics for biology and health*, Springer Science - Business Media, 2010, doi: 10.1007/978-1-4419-1742-3_10.
- [9] WHO, *Obesity: preventing and managing the global epidemic*, World Health Organization, Geneva, 2000.
- [10] WHO, *Waist circumference, and waist-hip ratio: Report of a WHO expert consultation*, World Health Organization, Switzerland, 2008.
- [11] S. Noora, "Neck circumference as an indicator of overweight and obesity in young adults," *American Journal of Applied Mathematics and Statistics*, vol.6(5), pp. 176-180, 2018, doi: 10.12691/ajams-6-5-1.
- [12] D. W. Hosmer, and S. Lemeshow, *Applied logistic regression, Second edition*, John Wiley & Sons, New York, 2000.
- [13] D. S. Young, *Handbook of regression methods*, CRC Press Taylor & Francis Group, Broken Sound Parkway, NW, Suite 300 Boca Raton, FL, 2017.
- [14] H.J. Motulsky, and A. Christopoulos, *Fitting models to biological data using linear and nonlinear regression. A practical guide to curve fitting*, GraphPad Software Inc., San Diego CA, 2003.
- [15] S.M.Pepe, *The statistical evaluation of medical tests for classification and prediction*, Oxford: Oxford University Press, 2003.
- [16] W. J. Krzanowski, and H. J. David , *Roc curves for continuous data*, CRC Press, Boca Raton, 2009.
- [17] M. Gonen, *Analyzing receiver operating characteristic curves with SAS*. SAS Institute Inc., Cary, NC, USA, 2007.