Binary Logistic Regression on Cafeteria Satisfaction Services

Norziha Che Him, Yusliandy Yusof, Nur Syazwani Aqilah Mohamad Aris

Abstract— This study investigates student's satisfaction level towards cafeteria services in Universiti Tun Hussein Onn Malaysia (UTHM) located in southeast Malaysia. A structured self-administered questionnaire survey has been conducted with 360 respondents by using stratified random sampling. This study adopted the Chi-Square test, Likelihood Ratio test and Binary Logistic Regression. The comparison result shows that students more satisfied to the Campus Cafeteria compared to the College Cafeteria. A significance test for the logistic coefficient by using the Likelihood Ratio test with predictors Food Quality, Staff Skills, Waiting Time and Gender show strong significant predictors that influenced student's satisfaction towards cafeteria services. Hosmer-Lemeshow test revealed the greater p-value of Model 1 (0.418) compared to Model 2 (0.261). Therefore, Model 1 has been chosen as the best model with Food Quality, Staff Skills, Waiting Time and Gender were significant factors in influencing the student's satisfaction towards the cafeteria.

Index Terms: Binary Logistic Regression, Cafeteria Satisfaction, Chi-Square Test Model Comparison

---

1. INTRODUCTION

Customer satisfaction is an accomplishment response and judgement on the service features or services that could provide a different level of consumption. It defined as a centre of marketing and the potential purpose to achieve customer's satisfaction and boost the quality of the serving meals [1]. It also known as the centre of marketing to achieve the customer's satisfaction [2]. Cafeteria services in Universiti Tun Hussein Onn Malaysia (UTHM) Pagoh Campus have two locations named as Campus Cafeteria and College Cafeteria. These cafeteria services have received various complains and reports from staff and students due to the lack of satisfaction services in many aspects during early establishment of cafeteria operation around 2018.

2. LITERATURE REVIEW

Previous study on the student's satisfaction towards the Café Giunto by using the multiple regression analysis by [3] shows an association exist between student's satisfaction and services quality. Therefore, it leads firmed direction that the service quality at the Café Giunto have met student's satisfaction. Similar study by [4] focused on the ambiance, food quality, price fairness and staff performance.

Meanwhile, their dissatisfaction escalated with the incompetent staff whose served at the counters, which had causes for a long waiting time and therefore long queue. These problems affect the comfortless of the other students that wish to buy a certain favourite food [5, 6]. This study aims to provide extra information to university administration in improving services and gain student's trust towards food services afterwards. Based on the several informal discussions and observation among the students, there are several matters and important issues were raised up about cafeterias services in UTHM. The main complaint is the inconsistent prices of foods at any time, even though students bought the same quantity of foods at the same times. Another problem is the menu design by campus cafeteria are not attractive and unclear to attract student to buy.

3. METHODOLOGY

Data was obtained through 360 self-administered questionnaires from the UTHM Pagoh Campus students through stratified samplings. Descriptive analysis, Chi-Square test and Binary Logistic Regression adopted in this study using R software. The Chi-Square test of independence or Pearson Chi-Square test is one of most useful statistics of hypothesis testing hypotheses used if variables are nominal. It could provide an information on the significance of any observed differences and on exactly which categories account for any differences found [7,16].

Generalised Linear Model (GLM) refer to a larger class of models popularized by [17]. GLM is a broad class of model that include a linear regression, Analysis of Variance (ANOVA), Poisson regression and log-linear model. It is an extension of linear model framework that included the range of response variable that is restricted and the variance of the response variable depend on the mean

- *Norziha Che Him is a Senior Lecturer in Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, 84600 Pagoh, Johor, Malaysia.*
  E-mail: norziha@uthm.edu.my
- *Yusliandy Yusof is a Master's Degree student in Faculty of Technical and Vocational Education, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, Malaysia.*
  E-mail: yusliandy@gmail.com
- *Nur Syazwani Aqilah Mohamad Aris is a Master's in Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, 84600 Pagoh, Johor, Malaysia.*
  E-mail: aqilahknowledge@gmail.com

value [8,9]. GLM is a development of the linear model to accommodate both non-Gaussian response distributions and transformations for non-linearity in the systematic model component.

The independent observations $(y_1,\ldots,y_n)$ on a response, where the distribution of y_i is in the exponential family with parameters θ_i and ϕ and with functions $a(\phi)$, $b(\theta_i)$, $c(y_i,\phi)$ chosen to be appropriate for the particular data as in (1).

$$p(y_i;\theta_i,\phi)=\exp\left[\frac{(y\theta_i-b(\theta_i))}{a(\phi)}+c(y,\phi)\right] \quad (1)$$

Here, there are explanatory variables (predictors), $x_i=(x_{i1},\ldots,x_{ip})$ whose values may influence the distribution of the response y_i through a linear predictor; $\eta_i=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\ldots+\beta_p x_{ip}$ which effects the mean of the response through smooth and invertible link function $g(\cdot)$ so that

$g(\mu_i)=\eta_i=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\ldots+\beta_p x_{ip}$ function.

Here, the link function $g(\cdot)$ does not transform y_i, but rather its mean μ_i. A Gaussian linear model with response $\log y_i$ is different with GLM with normal error and a log link [9,10].

Binary logistic regression estimates the probability that characteristic is present such as estimate probability of success given the values of explanatory variables. Let Y be a binary response variable, $Y_i=1$ if the trait is present in observation i and $Y_i=0$ if the trait is not present in observation i. Meanwhile, $X=(X_1, X_2, \cdots, X_k)$ be a set of explanatory variables which can be discrete, continuous or a combination.

Here, X_i is the observed value of the explanatory variables for observation i. Equation 2 and (3) show the basic models of binary logistic regression

$$\pi_i=\Pr\left(Y_i=1\mid X_i=x_i\right)=\frac{\exp\left(\sum_{k=1}^{p}\beta_k x_{ik}\right)}{1+\exp\left(\sum_{k=1}^{p}\beta_k x_{ik}\right)} \quad (2)$$

$$\text{logit}\,(\pi_i)=\log\left(\frac{\pi_i}{1-\pi_i}\right)=\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik} \quad (3)$$

Even though binary logistic regression has less assumptions compared to the GLM, it still requires other assumptions that should be follows before proceeding with the modelling [11]. First, binary logistic regression requires the response variable to be in dichotomous. Secondly, it is necessary to code the response variable accordingly with 1 as the desired outcome since binary logistic regression assumes that $\Pr(Y_i=1)$ is the probability of occurring event. Third, the model should be fitted accordingly and correctly by including all the meaningful variables into the model. Next, each of the observation should be independent and have a little or no multicollinearity.

Moreover, binary logistic regression requires linearity of the relationship between the explanatory variables and log odds. In the meantime, it does not require a linear relationship between the response variable and explanatory variables. Lastly, binary logistic regression requires large sample size since the maximum likelihood estimates are less powerful than Ordinary Least Square (OLS).

4. RESULT AND DISCUSSION

4.1. Contingency Tables and Chi-Square Test

The response variable is student satisfaction towards cafeteria services with coded value 1 as Satisfied and coded value 0 as Not Satisfied (see Table 1). Meanwhile, the explanatory variables are factors that influenced student's satisfaction towards cafeteria services which are gender, age, race, year of study, faculty, cafeteria location, take away or eat in, serving time, waiting time, seat, food quality, staff skills, ambience and price.

TABLE 1
DISTRIBUTION OF STUDENT'SS SATISFACTION AND LOCATION OF CAFETERIA

Cafeteria Location	Not Satisfied	Satisfied
Campus Cafeteria	40	140
College Cafeteria	80	100

Based on Table 1, the respondents are more satisfied with the services that provided by the Campus Cafeteria compared to the College Cafeteria with 140 respondents chosen satisfied with Campus Cafeteria, rather than only 100 respondents chosen satisfied with the College Cafeteria.

The Chi-Square test of independence based on contingency table in Table 1 that cafeteria location is significant and confirmed influenced the student's satisfaction at p-value less than 0.05.

4.2 Model Development

To determine an appropriate model for the Binary Logistic Regression, the explanatory variables were selected by using the Likelihood Ratio test and the variables with p-value more than 0.05 will be excluded from the model. There are two potential models, known as Model 1 with its explanatory variables were selected by using the

Likelihood Ratio test (see Table 3), meanwhile Model 2 is the model that based on the possible explanatory variables that may influenced the student's satisfaction as shown in Table 5.

TABLE 2
COEFFICIENTS AND EXPLANATORY VARIABLES FOR MODEL 1

Coefficient	Variables
α	Intercept
β_1	Gender
β_9	Waiting Time
β_{11}	Food Quality
β_{12}	Staff Skills

Model 1 consists factors that influenced student's satisfaction known as gender, waiting time, food quality and staff skills as explanatory variables for the model. In addition, Table 3 also shows an estimated coefficient, standard error, Z-value and *p*-value of Model 1 with all variables significance at least at < 0.100.

TABLE 3
ESTIMATED COEFFICIENTS, STANDARD ERROR, Z-VALUE AND *p*-VALUE OF MODEL 1

Coefficient	Estimate	Std. Error	Z-Value	*p*-Value
Intercept	-7.971	1.019	-7.817	< 0.001
β_1 (1)	1.000	0.455	2.198	< 0.050
β_9 (1)	1.004	0.303	3.318	< 0.001
β_9 (2)	-0.206	0.612	-0.337	< 0.100
β_{11}	2.190	0.377	5.816	< 0.001
β_{12}	1.204	0.307	3.922	< 0.001

Meanwhile, (4) is a logit equation for Model 1 where four explanatory variables show positive relationship with the student's satisfaction, which are Gender (female) [β_1], Waiting Time (less than 5 minutes) [β_9], Food Quality [β_{11}] and Staffs Skills [β_{12}].

$$\log\left(\frac{p}{1-p}\right) = -7.971 + 1.000\,\beta_1 + 1.004\,\beta_9 + 2.190\,\beta_{11} + 1.204\,\beta_{12} \quad (4)$$

TABLE 4
COEFFICIENTS AND EXPLANATORY VARIABLES FOR MODEL 2

Coefficient Variables	Coefficient Variables
α	Intercept
β_6	Cafeteria Location
β_{11}	Food Quality
β_{12}	Staff Skills
β_{13}	Ambience
β_{14}	Price

$$\log\left(\frac{p}{1-p}\right) = -7.837 + 0.804\,\beta_6 + 1.959\,\beta_{11} + 0.220\,\beta_{12} + 1.166\,\beta_{13} + 0.416\,\beta_{14} \quad (5)$$

TABLE 5
ESTIMATED COEFFICIENTS, STANDARD ERROR, Z-VALUE AND *p*-VALUE OF MODEL 2

Coefficient	Estimate	Std. Error	Z-Value	*p*-Value
Intercept	-7.837	1.032	-7.598	< 0.001
β_6 (1)	0.804	0.317	2.538	< 0.010
β_{11}	1.959	0.368	5.325	< 0.001
β_{12}	0.220	0.366	0.601	< 0.001
β_{13}	1.166	0.319	3.659	< 0.100
β_{14}	0.416	0.286	1.451	< 0.100

Meanwhile, (5) is a logit equation for Model 2 with five explanatory variables shows positive relationship with the student's satisfaction at *p*-value < 0.100, which are Cafeteria Location [β_6], Food Quality [β_{11}], Staff Skills [β_{12}], Ambience [β_{13}] and Price [β_{14}].

Table 4 presents five main factors that have been considered to be in the Model 2 which are Cafeteria Location, Food Quality, Staff Skills, Ambience and Price as explanatory variables. Therefore, Cafeteria Location, Food Quality, Staff Skills, Ambience and Price become main factors that influenced student's satisfaction towards the cafeteria services in UTHM Pagoh Campus.

4.3 Model Comparison

Deviance is a goodness-of-fit statistic for any statistical models that often used for statistical hypothesis testing. It's a generalization of the idea of using the sum of squares of residuals in ordinary least squares to cases where model-fitting is achieved by maximum likelihood and plays an important role in exponential dispersion models and GLM [13,14]. Meanwhile, Akaike information criterion (AIC) is an estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection [15].

In addition, Hosmer-Lemeshow (HL) test is a goodness-of-fit test for logistic regression where the test provided how well the data fits the model. HL test calculates if the observed event rates match the expected event rates in the population subgroups and has been used

for binary response variables. Basically, the output returns a chi-square value or known as a Hosmer-Lemeshow chi-squared and a *p*-value where the small *p*-value presents that the model is a poor fit [12]. Model comparison between Model 1 and Model 2 were based on the value of Deviance and AIC. Table 6 shows that Model 1 as the best model with the lowest AIC value compared to Model 2.

TABLE 6
DEVIANCE (D), PARAMETER (P), DF AND AIC

Model	Deviance, D	Parameter (p)	df	AIC
Model 1	458.29	5	355	323.19
Model 2	458.29	6	355	329.94

TABLE 7
HOSMER-LEMESHOW TEST

Model	df	Chi-Square	*p*-Value
Model 1	5	8.16	0.418
Model 2	6	10.06	0.261

Table 7 presents that Model 1 is said to be more fit to the data since the *p*-value is greater than the *p*-value of Model 2 and the nearest to approach to the value of 1. As a conclusion from the evaluation of the models based on Deviance, AIC and Hosmer-Lemeshow test, therefore Food Quality (β_{11}), Staff Skills (β_{12}), Waiting Time (β_9) and Gender (β_1) were found to be statistically significance to influenced student's satisfaction towards the cafeteria services in UTHM Pagoh Campus.

5. CONCLUSION

Conclusion could be made that students are more satisfied to the Campus Cafeteria compared to the College Cafeteria and the significance test for the logistic coefficient using Likelihood Ratio Test shows that Food Quality, Staff Skills, Waiting Time and Gender were significant factors in influenced the student's satisfaction towards the cafeteria. Hosmer-Lemeshow test was tested and the *p*-value of Model 1 is 0.418 and better than Model 2. Thus, Model 1 was chosen as the best model with the equation;

$$\log\left(\frac{p}{1-p}\right) = -7.971 + 1.000\beta_1 + 1.004\beta_9 + 2.190\beta_{11} + 1.204\beta_{12}$$

REFERENCES

[1] K. King and L. Harr, "Service Dimensions of Service Quality Impacting Customer Satisfaction of Fine Dining Restaurants in Singapore", Master Thesis, Bachelor of Arts, National University of Singapore, 2008.

[2] S.S. Andaleeb and C. Conway, Customer Satisfaction in the Restaurant Industry: An Examination of the Transaction-Specific Model, Journal of Services Marketing, 20(2), 3-11, 2006.

[3] A.A. Zainol, A.S. Kisun, S.N.I. Norashikin and Z. Mohi, "Students Satisfaction Towards Café Giunto", Thesis, UiTM Campus Puncak Alam, 2014.

[4] M.L. Dyana, M.S. Norazah and Y.L. Adeline, "Students Satisfaction with the Service Quality of Cafeteria: A Structural Approach", International Journal of Business, Economics and Law 4(1), 2014.

[5] A. Archakova, "Service Quality and Customer Satisfaction. Case study: Company X", Degree Thesis, Faculty of Business Administration, Saitama University of Applied Science, 2013.

[6] S.N. Dollah, N. Mansor and M. Mohamed, "Exploring the Major Determinants of Student Satisfaction on University Cafeteria Food Services: A Malaysian Case", Interdisciplinary Journal of Research in Business, 2(7), 62-73, 2012.

[7] M.L. McHugh, "The Chi-Square Test of Independence", Biochemia Medica, 23(2), 143-149, 2013.

[8] D.W. Hosmer and S. Lemeshow, "Applied Logistic Regression", 2nd ed, John Wiley & Sons, 2000.

[9] J.A. Nelder and R.W.M. Wedderburn, "Generalized Linear Models", Journal of the Royal Statistical Society, Series A (General), 135(3), 370-384, 1972.

[10] W.N. Venables and C.M. Dichmont, "GLMs, GAMs and GLMMs: An Overview of Theory for Applications in Fisheries Research", Fisheries Research, 70(2-3), 319-337, 2004.

[11] N. Torosyan, "Application of Binary Logistic regression in Credit Scoring", Master Thesis, Institute of Mathematics and Science, University of Tartu, 2017.

[12] D.W. Hosmer, S.A. Lemeshow and R.X. Sturdivant, "Applied Logistic Regression", 3rd ed. Hoboken, NJ: Wiley, 2013.

[13] B. Jorgensen, "The Theory of Dispersion Models, Chapman & Hall", 1997.

[14] P.X.K. Song, "Correlated Data Analysis: Modeling, Analytics, and Applications", Springer Series in Statistics, 2007.

[15] M.J. Brewer, A. Butler and S.L. Cooksley, "The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity", Methods in Ecology and Evolution, 7, 679-692, 2016.

[16] R. Miller and D. Siegmund, "Maximally selected Chi-square statistics", Biometrics, 38, 1101-1106, 1982.

[17] P. McCullagh and J.A. Nelder, "Generalized Linear Models", 2nd Edition, Chapman and Hall, London, 1989.