Volume: 2 Issues: 4 [June, 2017] pp.30-40]
International Journal of Accounting, Finance and Business
eISSN: 0128-1844
Journal homepage: www.ijafb.com

MULTINOMIAL LOGISTIC REGRESSIONS: FACTORS AND PREDICTION ON MALAYSIAN FILM VIEWERS

Noraini Abdullah[1], Diana Hasan[2], Suhaimi Salleh[3] & Zainodin H.J.[4]

[1]Senior Lecturer, Faculty of Science & Natural Resources, Universiti Malaysia Sabah
[2] Postgraduate, Centre of Postgraduate Studies, Universiti Malaysia Sabah, Malaysia
[3]Senior Lecturer, Faculty of Humanities, Arts & Heritage, Universiti Malaysia Sabah
[4] Co-Founder & President of Lex Capital Sdn Bhd. (1197516-D), Menara Maxis, KLCC

Accepted date: 28 May 2017, **Published date**: 5 July 2017

To cite this document:
Abdullah, N., Hasan, D., Salleh, S., & H.J., Z. (2017). Multinomial Logistic Regressions: Factors and Prediction on Malaysian Film Viewers. *International Journal of Accounting, Finance and Business (IJAFB), 2*(4), 30-40.

Abstract: *Many researchers are currently interested in conducting a technique in data analysis with qualitative dependent variables involving more than two categories, known as Multinomial Logistic Regressions (MLR). This study too had conducted using MLR to examine the significant factors based on the frequency of watching categories, such as 1) once or nil in a month, 2) twice a month, 3) three to four times in a month and 4) more than five times per month. From the frequency of watching categories (1, 2, 3 & 4), the frequency of watching more than five times per month (Cat 4) was referred to as the reference group, while the other categories had exhibited 255 MLR models each. Statistical tests, modelling procedures and models' goodness-of-fit tests were carried out on a total of 765 models from the 3 categories on film watching frequencies. In order to obtain a set of selected models (with significant variables), a progressive elimination (one by one, least significant first) of the insignificant variables was employed at Phase 2 of the model building procedures involving three types of tests namely NPC/NPM, multicollinearity and coefficient tests. Criteria based on pseudo R-squared were proposed consisting of Cox & Snell, Nagelkerke and McFadden, to finally single out the best model. The important findings highlighted in this study were the best model validation using the Mean Absolute Percentage Error (MAPE). Via the best models from each category, the model-building approach in Multinomial Logistic Regression analysis was established, and prediction using MAPE was done. Findings showed that the best models from all the respective categories (1, 2 & 3) had two common significant factors on the dependent variable. The results also showed that the best model from Cat 1 had the least MAPE (6.57%, thus indicated it was excellent to be used for prediction. Based on this, it is suggested that to attract more viewers, less films should be produced in a year, however, the allocated budget for film making should be focused on producing films which conformed to the identified significant factors that would attract more viewers. By using the best model, film viewing frequencies and number of film viewers can thus be predicted, and the expected revenue for the film industry can thus be estimated.*

Keywords: *Multinomial Logistic Regression (MLR), categories, significant factors, MAPE, best model.*

Introduction

Film being universal medium, has a unique cultural value. Besides, it is remarkably accessible and inclusive with its appeal traversing eras and intersection, across national and phonetic boundaries. In addition, it is able to confront people with the real world while speaking to its imagination, can be informative and revealing essential truths on human conditions and situations. Film also is accessible with immediacy, and when viewed at the cinema will provide an immersive experience. It has the ability to influence viewer's attitude and perception, especially to an avid viewer. The word "obsession" was etymologically used to portray the association with film. For avids, film viewing in cinema offers more than stories told in light and sound, seen once and soon forgotten. They are exceptionally visited by cinema goers that frequently go to film celebrations and seasons. They are attracted to free silver screens of which film is integral in their social life. When compared to non avid viewers who seldom go to the cinemas and usually based on their interest to a certain movie or influenced by only the box-office movies. It is common for avids to cite a particular film as the formative influence on their development. According to Salwen and Stacks (2006), the uses and gratifications theory stated that "people's media consumption patterns are intended actions on the part of the viewers" and that "individuals do make conscious choices about what they see and read in the media". This thus meant that audiences did not always choose to see a specific movie for the same reasons. Thus, this study is aiming to find out the most influential factors that would get the Malaysian film viewers to go to see a specific movie at the cinema This would ultimately contribute to the increase in the frequency of film viewing, and thus the number of film viewers.

Literature review

According to Miller (1999), film had the ability to influence viewers' perception because it provided information and pseudo-experiences, particularly in the absence of an individual's own experience. According to Kusumarasdyati (2004) and Luo (2004), film can catch the learners' interest, and it can positively affect students' motivation to learn. While, Mustafa (2009) had studied that seven factors had helped the Egyptian audiences in determining their choice of films such as movie stars, directors, trailers, general advertising, word of mouth, movie genre and reviews. Walmsley (2011) had explored the fundamental drivers behind theatre-going and to fill a gap in the literature on audience motivation. The study achieved a comprehensive qualitative study of theatre-going at the Melbourne Theatre Company and West Yorkshire Playhouse, which was carried out in 2010. The methods employed a combination of qualitative techniques, including responsive depth interviews and participant observation. The study found that the key motivation factors for participants were the pursuit of emotional experiences and impact.

Dyna (2012) stated that many researches had been done, but none of the researchers were able to develop parsimonious consumer decision making model that explained cinema decision making process involving many factors. Mohammadian and Habibi (2012) had also discovered that four influential factors had attracted the Iranians to go to the cinemas; they were namely, product, price, places and promotional factor. Besides that, Yang and Zhong

(2016) analyzed the effect of the perception of film attractiveness on the audience satisfaction, intention, investment and the mediating effect on satisfaction. The findings demonstrated that satisfaction mediated film attractiveness and intention. The satisfaction path of film attractiveness were perception of film attractiveness, investment; while mind purification, logic and inspiration were main factors to improve intention satisfaction.

According to Shelley and Allan (2016), Multinomial Logistic Regression (also referred to as polychotomous logistic regression) is frequently used for the analysis of categorical response data with continuous or categorical explanatory variables. Parameter estimates are usually obtained through direct maximum likelihood estimation. Janani and Umamaheswari (2014) also stated that Multinomial Logistic Regression can be generally used for solving problems that have multiple classes (i.e.) the dependent variable can be detected from a set of independent variables. The machine learning algorithm contains a set of ‘n’ training test to demonstrate a classifier with the given path of length ‘l’. Multinomial logistic recession uses a linear predictor function f (k, i) to forecast the prospect that the observation i has a consequence k.

Methodology

Modelling concepts of Multinomial Logistic Regression

In this study, mathematical modelling is employed to determine the factors that might influence viewers to go to cinemas. Logistic regression could be considered as a nonlinear regression (Kutner *et al.*, 2008). When there are only two categories of the dependent variable, multiple binary logistic (MBL) regression is regularly used rather than the discriminate analysis since there is a mixture of numerical and categorical independent variable(s). The logistic regression is easier to use because it includes procedures for generating the necessary dummy variables automatically, requires fewer assumptions, and is more statistically robust. It is also necessary when the independent variables are categorical or a mix of continuous and categorical, and the dependent variable is categorical. It forms the best fitting equation or function using the maximum likelihood method which maximizes the probability of classifying the observed data into the appropriate category, given the regression coefficients.The multiple binary logistic (MBL) regression can be used wisely, especially when the dependent variable is qualitative, and is used to predict the binary response when the dependent variable is dichotomous. According to Halcousis (2005), the logit model is based on the cumulative logistic regression, and it will give the probability estimates that are bounded by 0 and 1. However, in this study, the Multinomial Logistic Regression (MLR) is used instead to analyze the significant factors on movie viewing activities. The MLR model is a more appropriate model with regard to other regression models since the categorical dependent variable is nominal with more than two levels.

The general multinomial logistic regression model is given by: $Y_i = \Omega_0 + \Omega_1 W_1 + \Omega_2 W_2 + \ldots + \Omega_k W_k + u$ …(1) for j =1, 2, … k (Zainodin and Khuneswari, 2010) where ‘Y_i’ is the categorical dependent variable, ‘W_j’denotes the j-th variable, ‘Ω_0’ is the constant term of the model, ‘Ω_j’ is j-th coefficient of independent variable Wj, ‘k’ is the number of the single independent variables, (k+1) is the number of parameters, and ‘u’ is the error term. This study was conducted to examine the significant factors based on the categorical variable frequency of viewing, Y_i, with i=1, 2, 3 & 4.

Multinomial Logistic Regression Model-Building Procedures

Before any statistical analyses on the raw data were carried out, data preparations were done that involved the process of cleaning and organizing the data, as shown in Figure 1. According to Noraini *et al.* (2015), data cleaning was one of the prerequisites in statistical modelling to avoid biased and misinterpretation of the results. Next, factor analysis was used to examine the factors that influenced film audiences to become avid viewers. Factor analysis was an effective tool in reducing the dimensionality of a multivariate analysis (Bartholomew, 1980). Factors could be determined using factor analysis based on assumption that correlations were derived from scores that produced linear relationship (Child, 2006). Transformation into dummy variables of the independent variables was further performed before model-building procedures were carried out.

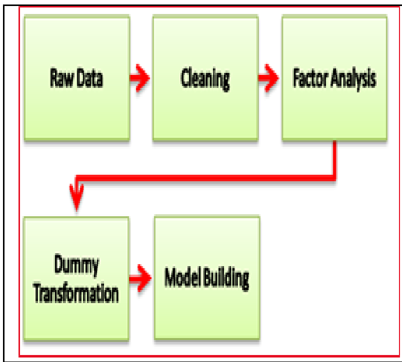
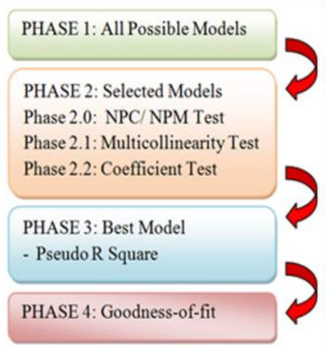


Figure 1: Data Preparation in Modeling Social Science Data



=>Number of all Possible models, $N=\sum_{j}^{q} {}^{q}C_j$

=>Model Labelling Ma.b.c
Ma = Full Parent Model
b = Variables removed in Multicollinearity Test
c = Variables eliminated in Coefficient Test

=>Best Model chosen based on the pseudo R-squared
=>GOF tests:Pearson, Deviance & Likelihood Ratio Tests.

Figure 2: Summary of Four-Phase Model-Building Procedures in MLR models

The Four-Phase Model-Building Procedures depicted in Figure 2 were carried out in this study. The multicollinearity test was based on the VIF approach (Zainodin *et al*., 2015). The best model chosen was based on the pseudo R-squared criteria, and finally the goodness-of-fit (GOF) tests were carried out on the best model to validity the model fitting for prediction and estimation. Further explanation and illustration on the modelling procedures can be referred in: (Zainodin *et al*., 2011; Noraini *et al*., 2016; Diana *et al*. 2017).

Results and Discussions

Data Collection and Preparation

Data sample were collected using questionnaires that focussed on the audiences that went to cinema in several states within Malaysia, namely Sabah, Sarawak, Johor, Selangor, Kedah and Pahang. These states were chosen due to the rapid increase in development and urban population, business centres equipped with many shopping malls and entertainment facilities provided like the cinemas. The questionnaire items were developed based on the research problem. The questionnaire comprised of two parts with 117 items. Part 1 of 23 items represented the demographic profiles of respondents, while part 2 represented the film viewer perceptions and factors to watch films. Significant demographic factors would include region, state, location, age, gender, ethnicity, religion, frequency of film viewing, education and salary. In this study, there were 1,337 respondents, comprised of 647 males and 688 females, and aged between 7 to 68 years. The raw data collected were in the field of social science with regard to the frequency of film viewing by cinema goers. Data cleaning using the row and column method was carried out on the raw data due to missing information from the respondents. After data cleaning, 1277 samples remained in the data set which thus had no missing values. Factor analysis was then performed as shown in Figure 1.

The scree plot as shown in Figure 3, plotted the eigenvalues against the number of components. The two components plotted above the red line, indicated that the eigenvalues had exceeded the value of 1.0. Table 1 displayed the rotated component matrix. Eight categorical independent variables from both components (1 & 2) that had the highest absolute value correlation greater than 0.50 (highlighted) were chosen as variables for model building procedures.

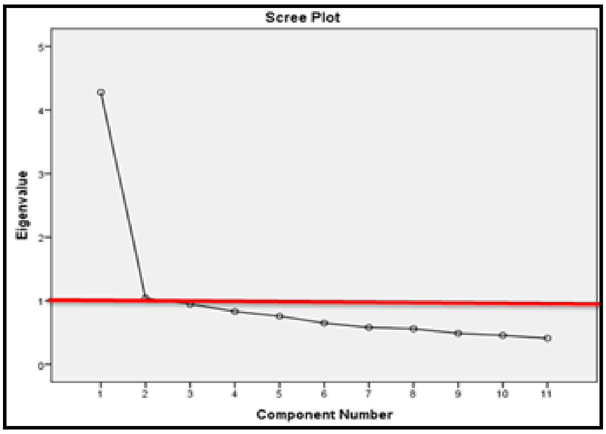


Figure 3:Scree Plot

Table 1: Rotated Component Matrix

	Component	
	1	2
Source of Information	.815	
Encouragement	.739	
Medium of Watching	.560	
Gratification	.442	.466
Film Production	.399	.301
Perception on Malay Film		.774
Public Perception		.635
Attraction	.374	.608
Theme	.395	.581
Genre	.490	.562
Showtime	.325	.478

Each category of the dependent variables was denoted as: Cat 1) once or never in a month, Cat 2) twice in a month, Cat 3) three to four times in a month, and Cat 4) more than five times per month. The category of more than five times per month (Cat 4) was referred as the reference group. Factors on viewing activities were i) Encouragement, ii) Source of information, iii) Gratification, iv) Film Genre, v) Perception on Malaysian films, vi) Medium of Watching, vii) Film production, and viii) Attractions of Watching. Data transformation were carried out on the 1277 samples, and data were further partitioned at 85% for modelling (n=1085), 10% for prediction (n=128), and 5% for estimating the missing values (n=64) using the best model. However, in this paper, only the modelling procedures with prediction using the Mean Average Prediction Error (MAPE) were illustrated. After factor analysis and data transformation, 255 models were obtained from each category (1, 2 and 3) as shown bythe formula given as in (2), and in Table 2 below

$$\boxed{N = \sum_{j=1}^{g} ({}^{g}C_j) = ({}^{8}C_1) + ({}^{8}C_2) + ({}^{8}C_3) + ({}^{8}C_4) + ({}^{8}C_5) + ({}^{8}C_6) + ({}^{8}C_7) + ({}^{8}C_8) = 255 \textit{ Models}} \quad .....(2)$$

Table 2: Phase 1 on All Possible Models for Each Category

Number of Independent Variables	1	2	3	4	5	6	7	8	Total	Models
Single Individual Variables	8	28	56	70	56	28	8	1	255	M1-M255

A total of 255 x 3=765 models from all the categories (1, 2 & 3) were modelled. For illustration purposes, model M13 from category 1 with variables (D_1, ..., G8) was chosen as shown in equation 3.

$$Y(\text{cat } 1) = f\,(D_1, D_2, D_3, D_4, D_5, G_1, G_2, G_3, G_4, G_5, G_6, G_7, G_8) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (3)$$

According to Zainodin *et al.* (2011), models with free from multicollinearity source variable and free from insignificant variable can be written as $M_{a.b.c}$, where 'm' denotes the model, 'a' denotes the number of the parent model, 'b' denotes the number of variables removed due to multicollinearity and 'c' denotes the number of variables eliminated due to variable insignificance. Phase 2.0 is the removal of the near perfect collinearity (NPC) and near-perfect multicollinearity (NPM) of the modelling procedures. Table 3 showed that Phase 2.0 of the NPC/NPM test on model M13 with all the independent variables had R^2 less than 0.95, thus no highly correlated variable was removed at this phase. Next, Table 3 also showed Phase 2.1 of the multicollinearity test which had all the VIF values less than 5.0. Hence, no elimination of multicollinearity source variable/s was carried out resulting in model M13 being denoted by model M13.0. Further illustrations can be referred in Diana *et al.* (2017).

Table 3: Phase 2.0 and 2.1 on Selected Model M13.0

PHASE 2.0

IV	R^2
D1	0.22122
D2	0.00921
D3	0.21014
D4	0.22402
D5	0.23529

PHASE 2.1

Coefficients		
Model	Collinearity Statistics	
	Tolerance	VIF
D1	.779	1.284
D2	.991	1.009
D3	.790	1.266
D4	.776	1.289
D5	.765	1.308

Table 4 below showed Phase 2.2 of the Coefficient test on the selected model M13.0. Two insignificant variables (D4, D3) with the highest p-value were eliminated due to insignificance, thus denoting the model as M13.0.2.

Table 4: Phase 2.2 Coefficient Test on Selected Model M13.0.2

Parameter Estimates

CAT 1ª	B	Std. Error	Wald	df	Sig.	Exp(B)	Interval for Exp(B) Lower Bound	Interval for Exp(B) Upper Bound
Intercept	.868	.145	36.005	1	.000			
[D1=0]	1.183	.210	31.729	1	.000	3.264	2.163	4.927
[D2=0]	.070	.153	.211	1	.646	1.073	.794	1.449
[D3=0]	.004	.172	.001	1	.980	1.004	.717	1.406
[D4=0]	-.003	.167	.000	1	.986	.997	.719	1.383
[D5=0]	.137	.195	.490	1	.484	1.146	.782	1.681

Parameter Estimates

CAT 1ª	B	Std. Error	Wald	df	Sig.	Exp(B)	Interval for Exp(B) Lower Bound	Interval for Exp(B) Upper Bound
Intercept	.867	.137	39.828	1	.000			
[D1=0]	1.183	.208	32.178	1	.000	3.263	2.168	4.910
[D2=0]	.071	.153	.211	1	.646	1.073	.794	1.449
[D3=0]	.003	.166	.000	1	.984	1.003	.725	1.388
[D5=0]	.136	.191	.507	1	.476	1.146	.788	1.665

Further insignificant variables were sequentially removed, until the variables that remained had p-values less than 0.05. Table 5 below showed further removals of variables D_2 and D_5 from the model until the p-value of the remaing variable was less than 0.05.

Table 5: Phase 2.2 Coefficient Test on Selected Model M13.0.10

Parameter Estimates

CAT 1ª	B	Std. Error	Wald	df	Sig.	Exp(B)	Interval for Exp(B)	
							Lower Bound	Upper Bound
Intercept	.868	.127	46.447	1	.000			
[D1=0]	1.183	.205	33.415	1	.000	3.266	2.186	4.878
[D2=0]	.070	.153	.212	1	.646	1.073	.795	1.447
[D5=0]	.137	.187	.534	1	.465	1.146	.795	1.654

Parameter Estimates

CAT 1ª	B	Std. Error	Wald	df	Sig.	Exp(B)	Interval for Exp(B)	
							Lower Bound	Upper Bound
Intercept	.911	.088	106.249	1	.000			
[D1=0]	1.188	.205	33.669	1	.000	3.279	2.195	4.897
[D5=0]	.134	.187	.515	1	.473	1.144	.793	1.650

Parameter Estimates

CAT 1ª	B	Std. Error	Wald	df	Sig.	Exp(B)	Interval for Exp(B)	
							Lower Bound	Upper Bound
Intercept	.934	.083	127.689	1	.000			
[D1=0]	1.239	.192	41.574	1	.000	3.453	2.369	5.032

Subsequent removal processes of insignificant variables were performed until all the variables that remained in the model had their p-values less than 0.05. The resulting model M13.0.10 was obtained after the zero elimination of the multicollinearity test, and the removal of 10 insignificant variables from the Coefficient test. Table 6 showed that Category 1 had thus obtained 42 selected models with the corresponding parameters (k+1) and total number of parameters (NP) before proceeding to Phase 3 of the model-building procedures.

Table 6: The corresponding selection criteria based on Pseudo R-squared of Category 1

NP	Model	Selected Model	(k+1)	*Cox and Snell R^2*	*Nagelkerke R^2*	*McFadden R^2*
6	M1	M1.0.4	2	0.0450	0.0689	0.0435
6	M2	M2.0.4	2	0.0078	0.0119	0.0074
↓	↓	↓	↓	↓	↓	↓
12	M15	M15.0.9	3	0.0562	0.0861	0.0546
↓	↓	↓	↓	↓	↓	↓
17	M75	M75.0.9	8	0.0777	0.1189	0.0763
↓	↓	↓	↓	↓	↓	0.0951
22	M100	M100.0.15	7	0.1066	0.1632	0.1064
↓	↓	↓	↓	↓	↓	↓
23	M134	M134.0.14	9	0.0616	0.0942	0.0600
↓	↓	↓	↓	↓	↓	↓
29	M175	M175.0.21	8	0.0893	0.1368	0.0884
↓	↓	↓	↓	↓	↓	↓
29	M200	M200.0.19	10	0.0743	0.1137	0.0729
↓	↓	↓	↓	↓	↓	↓
36	M240	M240.0.23	13	0.1105	0.1692	0.1106
41	M247	M247.0.28	13	0.1364	0.2088	0.1385
42	M253	M253.0.30	12	0.1172	0.1793	0.1176
			Maximum	0.1364	0.2088	0.1385

The best model was chosen based on the model that having the majority maximum value of pseudo R square, namely, criteria based on Cox & Snell, Nagelkerke and McFadden. Results in Table 6 showed that model M247.0.28 was chosen as the best model from category 1 where it had the maximum values of the pseudo R-square criteria. The factors of the best models from Category 2 and 3 respectively were shown in Table 7 below.

Table 7: Best Models From Category 1, 2 and 3

Category	Best Model From All Categories
1	M247.0.28 : Y_1= f(D_1, S_4, M_1, M_2, M_3, K_5, P_2, P_4, P_6, G_1, T_1, T_5)
2	M250.0.26 : Y_2= f(D_1, D_3, D_5, S_3, S_4, S_5, M_1, M_2, M_3, K_5, G_6, G_8, T_5, F_3)
3	M14.0.6 : Y_3= f(D_1, T_1, T_3, T_5)

The common significant factors of the MLR best models (Table 7- highlighted in green) were from (D_1), and (T_5) respectively. These variables were the most significant because both variables were selected as significant variables for all categories after going through the entire statistical tests and modelling procedures. Other variables that had significantly contributing in two categories would include (S_4), (M_1), (M_2), (M_3), (K_5), and (T_1) respectively (highlighted in yellow).

Table 8 depicted the description of the significant factors that had remained in the best models from Cat 1, Cat 2 and Cat 3 respectively. From Table 8, it could be said that frequencies on film viewing were most significantly affected by the categorical factors on Encouragement-From Interest and Attractions of Watching-From Friends/People Influence, followed by other significantly contributing factors which were source of information from newspapers and magazines, through advertisements in cinemas, internet and television, besides atraction of viewers from film's theme and gratification due to cultural or historical aspects of films. These can be seen from the increase in the number of film viewers to the cinemas with the influx of Hindi and Korean films.

Table 8: Description of Significant Factors for Each Category

Category	**Best Model**	**Significant Factor**
1	M247.0.28	D_1 = Encouragement - Interest S_4 = Source of Information – Newspaper/Magazine M_1 = Medium - Cinema M_2 = Medium - Internet M_3 = Medium - Television K_5 = Gratification- Culture/ History P_2 = Production of Film-Hindi / Tamil P_4 = Production of Film-Japan / Korea P_6 = Production of Film-Malaysia G_1 = Genre-Action T_1 = Attractions of viewer-Theme T_5 = Attractions of viewer-Influence Friends/People
2	M250.0.26	D_1 = Encouragement - Interest D_3 = Encouragement - Friends D_5 = Encouragement - Routine S_3 = Source of Information – Poster / Brochure S_4 = Source of Information – Newspaper/Magazine S_5 = Source of Information – Story from Friends

		M_1 = Medium - Cinema M_2 = Medium - Internet M_3 = Medium - Television K_5 = Gratification- Culture/ History G_6 = Genre – Science Fiksyen G_8 = Genre - Horror T_5 = Attractions of viewer-Influence Friends/People F_3 = Opinion movies Malay-Script
3	M14.0.6	D_1 = Encouragement - Interest T_1 = Attractions of viewer-Theme T_3 = Attractions of viewer-Actor T_5 = Attractionsof viewer- Influence Friends/ People

Goodness-of-fit (GOF) was being carried out to examine the goodness or appropriateness of the best model in fitting the data. The Pearson and Deviance Test were used to check the goodness-of-fit for the best model M247.0.28 from category 1. The hypothesis for Pearson and Deviance Test for best model M247.0.28 is as follows:

H_0: Model appropriate with data
H_1: Model not appropriate with data

The Deviance statistics (670.1612) test carried out with significant p-value of 0.1837 ($p>0.05$) had shown that model M247.0.28 is an appropriate model. The likelihood ratio test carried out on all the factors in the best model also showed that they were significant with p-values less than 0.05. Finally, the MAPE of each best model from the 3 categories were calculated and compared. It can be seen below that model from category 1 is the best to be used for prediction of the film viewers.

Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error or MAPE is called a justification for the accuracy of the model, which measures the average absolute errors by using the reserved data and the corresponding value of variables in the best model. When the MAPE equal to net zero, indicating that the forecast value is perfect fit to the actual value. According to the Panneerselvam (2004), MAPE is the average of the deviation of forecast demand and actual demand in term of percentage. Hence, the formula to calculate the MAPE as follows:

$$\text{MAPE (\%)} = \frac{1}{m}\left(\sum\left|\frac{A_t - F_t}{A_t}\right|\right) x\ 100$$

where, m is the number of observations in reserved data, A_t is the actual value of Y, and F_t is the estimated value of $\hat{Y}$ using the obtained best model. According to Juan *et al.* (2013),

MAPE has important and desirable features including reliability, unit-free measure, ease of interpretation, clarity of presentation, support of statistical evaluation and the use of all the information concerning the error, while in Zainodin *et al.* (2011), the reliability of best model can be achieved when the MAPE is smaller or equal to 10% with acceptability when MAPE is less than 25%.

The MAPE for each category were computed as below:

$$\text{MAPE (\%): category 1} = \frac{1}{128}\left(\sum\left|\frac{A_t - F_t}{A_t}\right|\right) = \frac{8.4109}{128} = 0.0657 = 6.57\%$$

$$\text{MAPE (\%): category 2} = \frac{1}{128}\left(\sum\left|\frac{A_t - F_t}{A_t}\right|\right) = \frac{11.0285}{128} = 0.0862 = 8.62\%$$

$$\text{MAPE (\%): category 3} = \frac{1}{128}\left(\sum\left|\frac{A_t - F_t}{A_t}\right|\right) = \frac{30.4706}{128} = 0.2381 = 23.81\%$$

Table 8: Summary of MAPE for each categories

Model	Category	MAPE (%)	Remark
M247.0.28	Cat 1	6.57%	Reliability Achieved
M250.0.26	Cat 2	8.62%	Reliability Achieved
M14.0.6	Cat 3	23.81%	Acceptable

Conclusion

Mathematical modelling concepts and procedures using multinomial logistic regressions (MLR) had determined the significant factors that influenced film viewers via their frequencies to the cinemas. These significant factors were found to have given positive and direct contribution to the increase in the frequency of film watching from encouragement-interest category and attraction of watching-influence from friend/people's category. This finding implied that more encouragement due to interest and attraction from friends or people would lead to higher frequency of viewers, more impact though to avids in going to cinemas, indirectly, to the number of viewers. In addition, this study also found that the source of information from newspapers and magazines, through advertisements in cinemas, internet and television, besides genre of theme and action with the influx of Hindi and Korean films had contributed film viewers to cinemas. Thus, it can be suggested to Malaysian film producers to focus on these factors towards producing high quality films of the Malaysian context, hence, further predicting the number of viewers and expected revenues from the film industry.

Acknowledgements

This study was partially supported by FINAS and the Centre of Research and Innovation of Universiti Malaysia Sabah under the grant number (SBK0249-SG-2015). Their contributions were gratefully acknowledged.

References

Bartholomew, D. J. (1980). Factor Analysis for Categorical Data. *Journal of the Royal Statistical Society: Series B*, 42 (3): 293-321

Child, D. (2006). *The Essentials of Factor Analysis.* 3rd Edition. New York: Cambridge: Cambridge University Press.

Diana Hassan, Noraini Abdullah, Zainodin H.J. and Suhaimi Salleh. (2017). Journal of Global Business and Social Entrepreneurship (GBSE), 1(4): 24-34. eISSN:24621714.

Dyna, H. S. (2012). Identifying Key Factors Affecting Consumer Decision Making Behavior in Cinema Context: A Qualitative Approach. *International Conference on Business, Economi, Management and Behavioral Science (ICBEMBS'2012)*. Dubai.

Halcousis, D. (2005). *Understanding Econometrics.* New York-South Western.

Janani P. R and Umamaheswari, K. (2014). Type 2 Diabetes Prediction Using Multinomial Logistic Regression. *Australian Journal of Basic and Applied Science,* 8(10): 31-37

Juan J.M.M., Alfonso, P.P., Albert, S.A. and Berta , C.B. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema,* **25** (4): 500-506.

Kusumarasdyati. (2004). Listering, Viewing and Imagination: Films in EFL Classes. The 2nd International Conference on Imagination and Education Vancouver, Canada, July 14-17, 2004.

Kutner, M. H., Nachtsheim, C.J. & Neter, J. (2008). *Applied Linear Regression Models.* 4th Edition. Singapore: McGraw-Hill Companies, Inc.

Luo, J.J. (2004). Using DVD films to enhance college freshmen's English listening Comprehension and Motivation. Unpublished Master thesis, National Tsing Hua University, Hsinchu. Taiwan.

Miller, K. S. (1999). Publication Relation in Film and Fiction: 1930 to 1995. *Journal of Public Relations Research,*11(1): 3-28

Mohammadian and Habibi. (2012). The Impact of Marketing Mix on Attracting Audiences to the Cinema. *International Business and Management,* 5(1): 99-106.

Mustafa. (2009). Going to the Movies: An Investigation of Factors Influencing Egyptian Audiences' Movie Choices. *Off Screen.* 13(6).

Noraini A., Liew L.K., Tan W.H., Zainodin H.J. and N.Surugau. (2016). Modelling Procedures in Determining Heavy Metals Concentration:A Case Study Using Barks of the Cinnamom Tree. International Journal of Mathematics and Computer Research, 4(6):1502-1513.

Noraini, A., Suhaimi, S., Zainodin, H. J. and Diana, H. (2015). Factors Relating Behavioral Patterns on Film Viewers. *Proceedings of INTCESS15-2nd International Conference on Education and Social Science,* Istanbul: 430-436

Panneerselvam, R. (2004). *Research Methodology.* New Delhi: Prentice-Hall of India Private Limited

Salwen, M. and Stacks, W. (2006). An Integrated Approach to Communication Theory and Research Lawrence Erilbaum Associates, Inc.

Shelley B. B & Allan, D. (2016). The Efficiency of Multinomial Logistic Regression Compared With Multiple Group Discriminant Analysis. *Journal of the American Statistical Association,* 82(400):1118-1122

Walmsley, B.A. (2011). Why People Go To The Theatre: A Qualitative Study Of Audience Motivation. *Journal of Customer Behaviour,* 10(4): 335-351. ISSN 1475-3928

Yang, D and Zhong, X. (2016). The Perception of Film Attractiveness and Its Effect on the Audience Satisfaction, Intentation and Investment. *Journal of Service Science*, 9: 21-27

Zainodin, H. J. and Khuneswari, G. (2010). Model Building Approach in Multiple Binary Logit Model for Coronary Heart Disease. *Malaysian Journal of Mathematical, Science,* 4(1): 109-136

Zainodin, H. J., Yap, S. J., and Noraini, A. (2011). An Alternative Multicollinearity Approach in Solving Multiple Regression Problem. *Trends in Applied Sciences Research*, **6**(11): 1241-1255

Zainodin, H. J., Khuneswari, G., Noraini, A. and Haider F. A. A. (2015). Selected Model Systematic Sequence via Variance Inflationary Factor. *International Journal of Applied Mathematics*(IJAPM),5(2):105-114.(ISSN:2010-362X).