

**MATHEMATICAL MODEL FOR BUYING BEHAVIOR  
OF INTERNATIONAL TRAVELLERS – A  
MULTINOMIAL LOGISTIC REGRESSION APPROACH**

Borukgamage Suravi Malinga Borukgama

(148853R)

Degree of Master of Science

Department of Mathematics

University of Moratuwa

Sri Lanka

July 2018

## DECLARATION

I hereby declare that this submission is my own work and that to the best of my knowledge and belief, it contains neither materials published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma or university or other institute of higher studies, except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

.....

B. S. Malinga Borukgama Date

Department of Mathematics

University of Moratuwa

The above candidate has carried out research for the Master's Dissertation under my supervision.

Name of the Supervisor: Prof. T. S. G. Pieris

.....

Signature of the supervisor Date

## **Abstract**

Predicting the future sales is required continuous attention in order to fulfill the consumer requirements. Disaggregating the sales in to micro level of the business would increase the prediction accuracy since the ground level requirements, trends and patterns are captured. Multinomial logistic regression is a technique that is to be used to model the outcomes with categorical response variable with more than two levels. In the study, the significant determinants for the brand wise purchasing decision of chocolates in a travel retail chain and its consequences are investigated. Multinomial logistics regression found that nationality of the consumers, time of purchase, preference for promotions and preference for weight of the products have significant impact on the chocolate brand choice. It was also found that these fours variables have no multicollinearity effect. The pseudo R<sup>2</sup> value of the model confirms that only 44.9% of the variability is absorbed by the final model. The model has overall brand classification accuracy of 52.4. The buying preference for any brand of cholate is maximized during the 1st quarters. Mix and match promotion maximizes the preference for purchase of Mars and Mondelez brands while buy 3 get 1 free become the promotion that maximizes the buying preference for Nestle. Preference for weight category is variant for the 3 brands. The relative nationality wise probabilities of selecting a brand of chocolate for fixed levels of promotional preferences and product weight preferences are derived with the multinomial logistic transformation equations. When comparing the nationality wise brand selection probabilities, no significant changes to the probabilities were found according to the nationalities of the customers. It is recommending to carry out similar studies for other sales as well.

**Keywords:** Chocolate Purchase, International Travelers, Likelihood Ratio Test, Multicollinearity, Multinomial Logistic Regression,

**DEDICATION....**

.... May this research be dedicated to my ever-loving parents, Wife  
Indeewari and Daughter Sasmini .....

## **ACKNOWLEDGEMENT**

I would make this an opportunity to give my heartiest gratitude to everyone who put their maximum effort to encourage me to complete this study in a successful way.

My initial gratitude goes to the Head of the Department of Mathematics and all other resource persons during the Masters' Degree program who guides me to improve the knowledge. A special thank goes to Professor T. S. G. Pieris, Professor in Applied Statistics for supervising and guidance given until the task is completed. Guidance and timely feedbacks given with related to the theoretical background would really helped me on improving the desire to search and find the solution to the research problem.

I should give my gratitude to Mr. Samer Orbiete, General Manager World Duty Free Group Lanka Limited and Mr. Dammith Amarasiri, Head of Supply Chain for facilitating me the data collection. Ms. Razeena Rifae, Assistant Manager Supply Chain helped me on giving insights on data collection and knowledge on technical background regarding the forecasting at operational level. My special gratitude goes to Mr. Sidath Perera – Head of Tea Buying, Blending and Exports for releasing me for the academic activities as and when required. Further, I would like to thank Ms. N. D. Namarathna, Controller, IT and Operations, Department of Immigration and Emigration for granting me the approval to collect the airline passenger movement data. Mr. Saliya Caldera was corporative on providing the data in user friendly formats.

Further I would like to extend my thanks to my colleagues in the Masters' Degree in Business Statistics (2014 / 2015) batch as well as the bachelor's degree batch (2006 / 2007), especially Malith Amarsinghe, Manoj Rodrogo, Gayathri Liyanage and Shirekha Layangani.

Finally, I would like to thank to my Amma, Thaththa, Ayya and Akka for their encouragement and loving support during the hard times. A Special thank goes to my wife, Indeewari for her enormous encouragements and support given during the entire period of research work. Some of her feedbacks and ideas gave me fresh avenues to continue when there were dead ends. And finally, to all the others who helped me in various ways to compete this project report.

## TABLE OF CONTENTS

DECLARATION .....	i
Abstract .....	ii
Dedication .....	iii
Acknowledgement .....	iv
List of Figures .....	viii
List of Tables .....	ix
List of Equations .....	xi
List of Abbreviations .....	xii
<b>1 INTRODUCTION .....</b>	<b>1</b>
1.1 General Background .....	1
1.2 Theoretical Background .....	2
1.3 Research Problem .....	3
1.4 Objectives .....	5
1.5 Limitations .....	6
1.6 Outline of the Dissertation .....	6
<b>2 LITERATURE REVIEW .....</b>	<b>8</b>
2.1 Qualitative Studies on Buying Behaviors .....	8
2.2 Studies on Buying Behaviors Regarding Food .....	9
2.3 Applicability of Different Forecasting Techniques .....	11
2.4 Aggregate versus Disaggregate Sales Forecasting.....	13
2.5 Discrete Choice Model .....	15
2.6 Applicability of Multinomial Logistic Regression Model .....	18

<b>3</b>	<b>MATERIALS AND METHODS</b> .....	<b>22</b>
3.1	Data Collection .....	22
3.1.1	Data Filtering and Preparation .....	23
3.2	Identify the Key Product Segments in Chocolate .....	23
3.3	Brand Choice (Yi).....	24
3.4	Factors influential for Chocolate purchasing Decisions .....	25
3.4.1	Nationality of the Passengers (X1) .....	25
3.4.2	Time of Purchase (X2) .....	29
3.4.3	Preference for Promotions (X3) .....	30
3.4.4	Preference for Product Weight (X4) .....	32
3.5	Data Preparation for Logistic Regression .....	33
3.5.1	Basic Statistics of Brand Choice Data.....	34
3.6	Theoretical Background.....	36
3.7	Odds .....	37
3.7.1	Odd Ratio .....	37
3.7.2	Difference of Log Odds.....	38
3.8	Logistic Regression Model .....	39
3.8.1	Logistic Regression and Discriminant Analysis .....	40
3.8.2	Logistic Curve .....	40
3.8.3	Logistic Transformation .....	41
3.8.4	Interpretation of Multinomial Regression Coefficient .....	43
3.8.5	Model Selection Problem .....	44
3.8.6	Specify Model in Custom Model .....	45
3.8.7	Significance of Explanatory Variables.....	45
3.8.8	Model acceptance criteria.....	47
3.8.9	The Classification Table.....	48
3.8.10	Model Diagnostic Tests .....	48
3.9	Steps of the Model Development.....	49
3.9.1	Identifying the Target Group.....	49
3.9.2	Definition of the Response Variable and Explanatory Variables.....	49
3.9.3	Two – way Contingency Tables .....	49
3.9.4	Derive the Multinomial Logistic Model.....	50
3.9.5	Calculate the Probability of Brand Selection .....	50

4	RESULTS AND DISCUSSION .....	51
4.1	Impact of Explanatory Variables on Brand Preference .....	51
4.1.1	Impact of Nationality on Brand Preference.....	51
4.1.2	Impact of Time of purchase on Brand Preference.....	53
4.1.3	Impact of Preference for Promotions on Brand Preference.....	54
4.1.4	Impact of Preference for Weight Category on Brand Preference.....	56
4.1.5	Correlation between Explanatory variables and Response Variable.....	57
4.2	Develop a Multinomial Logistic Model.....	58
4.2.1	Multinomial Logistic Regression Model.....	60
4.2.2	Significance of Effects .....	61
4.2.3	Goodness of Fit Test.....	62
4.2.4	Results of Pseudo R – Square.....	62
4.2.5	Classification Table .....	63
4.3	Interpretation of Parameter Estimates.....	64
4.3.1	Mars Relative to Other Brands .....	64
4.3.2	Mondelez relative to Other brands .....	68
4.3.3	Nestle respective to Other Brands .....	72
4.4	Probability of Brand Selection.....	76
4.4.1	Calculate the Brand Choice Probabilities.....	78
4.4.2	Nationality Impact on Brand Choice Probability .....	84
4.5	Summary of the Results .....	87
5	CONCLUSIONS AND RECOMMENDATIONS .....	88
5.1	Conclusions.....	88
5.2	Recommendations.....	88
	REFERENCES .....	90



**LIST OF FIGURES**

Figure 3-1 Logistic Curve ..... 41

## LIST OF TABLES

Table 3-1 Product segment wise volume and value - 2014 Jan to 2016 Dec.....	24
Table 3-2 Brand wise volume contribution for Sharing Packs - Jan 2014 to Dec 2016 .....	25
Table 3-3 Nationality wise total Departures - Jan 2010 to Dec 2016.....	26
Table 3-4 Levels of the nationality group.....	28
Table 3-5 Region wise Departures and Chocolate Purchase volumes - Jan 2014 to Dec 2016.....	29
Table 3-6 : Nationality wise Correlation between purchasing quantities and passenger outflow .....	29
Table 3-7 Levels of the time of purchase.....	30
Table 3-8 Levels of the Promotion Type .....	32
Table 3-9 Definition of the Sharing Requirement based on the product volume group .....	33
Table 3-10 Levels of the Sharing Requirement (per unit volume) .....	33
Table 3-11 Case Processing Summary.....	36
Table 4-1 Relationship between Nationality vs. Brand Preference .....	51
Table 4-2 Cross Tabulation - Nationality vs. Brand Preference .....	52
Table 4-3 Relationship between Time of Purchase vs. Brand Preference .....	53
Table 4-4 Cross Tabulation – Time of Purchase vs. Brand Preference .....	54
Table 4-5 Relationship between Preference for Promotion vs. Brand Preference..	55
Table 4-6 Cross Tabulation – Preference for Promotions vs. Brand Preference ...	56
Table 4-7 Relationship between Preference for Weight Category vs. Brand Preference .....	56
Table 4-8 Cross Tabulation – Preference for Product Weight vs. Brand Preference .....	57
Table 4-9 Correlation between response variable and explanatory variables.....	58
Table 4-10 Significance of the Explanatory Variables in Different Model Development Methods.....	59
Table 4-11 Step summary of the Backward Elimination Method.....	60
Table 4-12 Model Fitting Information of the Entry Method.....	60

Table 4-13 Effect wise likelihood ratio test results for Backward Elimination method .....	62
Table 4-14 Results of the Goodness of Fit test – Backward Elimination Method....	62
Table 4-15 Results for Pseudo R - Square – Backward Elimination Method.....	63
Table 4-16 Classification Accuracy of the Entry Model .....	64
Table 4-17 Parameter Estimates for Mars relative to other brands.....	66
Table 4-18 Parameter Estimates for Mondelez relative to other brands.....	71
Table 4-19 Parameter Estimates for Nestle relative to other brands.....	75
Table 4-20 Brand Preference Probability - fixed Weight Category 1 and promo category 1 .....	79
Table 4-21 Brand Preference Probability - fixed Weight Category 1 and promo category 2.....	80
Table 4-22 Brand Preference Probability - fixed Weight Category 1 and promo category 3.....	81
Table 4-23 Brand Preference Probability - fixed Weight Category 1 and promo category 4.....	82
Table 4-24 Brand Preference Probability - fixed Weight Category 1 and promo category 5.....	83
Table 4-25 Nationality impact on Brand Selection Probabilities – 2+1 Promotion activated .....	84
Table 4-26 Nationality impact on Brand Selection Probabilities – 3+1 Promotion activated .....	85
Table 4-27 Nationality impact on Brand Selection Probabilities – Dollar off Promotion activated .....	85
Table 4-28 Nationality impact on Brand Selection Probabilities – Mix and Match Promotion activated .....	86
Table 4-29 Nationality impact on Brand Selection Probabilities – No Promotions activated .....	86

## LIST OF EQUATIONS

Equation 2-1 Discrete choice model .....	17
Equation 3-1 Odds .....	37
Equation 3-2 Odd Ratio .....	38
Equation 3-3 Logistic Function.....	40
Equation 3-4 Logit Transformation .....	41
Equation 3-5 Logistic Transformation .....	42
Equation 3-6 General form of Multinomial Logistic Regression .....	43
Equation 3-7 General form of Multinomial Logistic Transformation .....	43
Equation 3-8 Calculate the parameter estimate of Logistic regression.....	44
Equation 3-9 Likelihood Ratio.....	46
Equation 3-10 Cox and Snell's R - Square .....	47
Equation 3-11 Nagelkerke's R-square .....	48
Equation 4-1 Multinomial Logistic Regression for selection of Mars.....	76
Equation 4-2 Multinomial Logistic Regression for selection of Mondelez.....	77
Equation 4-3 Multinomial Logistic Regression for selection of Nestle.....	77
Equation 4-4 Multinomial Logistic Transformation of selection of Mars.....	77
Equation 4-5 Multinomial Logistic Transformation of selection of Mondelez .....	77
Equation 4-6 Multinomial Logistic Transformation of selection of Nestle.....	78
Equation 4-7 Multinomial Logistic Transformation of selection of Other brands ..	78

## LIST OF ABBREVIATIONS

Abbreviation	Description
ANN	Artificial Neural Network
ARIMA	Auto – regressive Integrated Moving Average
BPNN	Back – Propagation Neural Network
BIA	Bandaranaike International Airport
CF	Sharing packs – over 175 grams’
MIS	Management Information System
MLR	Multiple Linear Regression
MSE	Mean Square of Error
MSR	Mean Regression Sums of Square
OLS	Ordinary Least Square
POS	Point of Sales
SKU	Stock Keeping Unit
SSR	Regression Sums of Square
SST	Total Sums of Square
USD	United States Dollars
VIF	Variance Inflation Factor