Bulletin of the *Transilvania* University of Braşov
Series V: Economic Sciences • Vol. 8 (57) No. 2 - 2015

Using the Logistic Regression model in supporting decisions of establishing marketing strategies

Cristinel CONSTANTIN[1]

Abstract: *This paper is about an instrumental research regarding the using of Logistic Regression model for data analysis in marketing research. The decision makers inside different organisation need relevant information to support their decisions regarding the marketing strategies. The data provided by marketing research could be computed in various ways but the multivariate data analysis models can enhance the utility of the information. Among these models we can find the Logistic Regression model, which is used for dichotomous variables. Our research is based on explanation the utility of this model and interpretation of the resulted information in order to help practitioners and researchers to use it in their future investigations.*

Key-words: *marketing strategies, multivariate data analysis, logistic regression, information, statistical significance*

1. Introduction

Every company needs to establish goals on short, medium and long term, which guide its overall activity toward success. While goals indicate what a business unit wants to achieve, the strategy design the ways for achieving these goals (Kotler & Keller 2006, 56). The marketing strategies establish the main actions that a company should put in practice on medium and long terms in order to achieve some specific goals like: market share, sales volume, profit ratio etc. (Florescu et al. 2003, 669). The marketing strategies are meant to reduce the risks in the market activity and to serve better the target segments taking into account the needs of these segments' members (Lefter et al. 2006, 687).

In establishing marketing strategies the managers have to define those groups and needs that the market offerings are intended to satisfy and then to establish specific actions to assure purchasing, manufacturing, sales, finance, and human resources meant to support the goals' achieving (Kotler & Keller 2006, 60). Strategic marketing also includes functions like: analysis of internal and external environment of the company, setting goals and marketing strategies, target markets,

[1] Transilvania University of Braşov, cristinel.constantin@unitbv.ro

positioning, product design, pricing, establishing distribution channels, establishing a communication strategy and promotion (Todor 2014, 61). Strategic marketing has crossed the border of goods and services manufacturing and it is used in various other social and economic fields such as city branding (Ispas at al. 2015, 19), sport, health, religion etc. But all these strategies have to be based on proper marketing researches meant to identify the needs, behaviours and attitudes of the targeted market segments. In this respect not only the methods of data collection are important but also the data analysis methods, which could be univariate, bivariate or multivariate methods.

The present paper is a part of a series of instrumental researches meant to review the main multivariate data analysis models that could help researchers and practitioners in interpreting the results of various marketing researches. The main advantage of the multivariate data analysis methods consists in the possibility to concentrate the simultaneous dependence or interdependence of many variables that can explain better some behaviours, attitudes and intentions. The results of these analyses can help decision makers in designing proper strategies and targeting market segments for these strategies.

Among the multivariate data analysis methods the Logistic Regression model is very useful due to the fact that it uses dichotomous variables that are often meet in the surveys' questionnaires. Such variables are used mainly for descriptive analyses or bivariate inferential analyses but they cannot be used in many multivariate models. Thus the Logistic Regression model can provide additional information extracted from the simultaneous effect of several dichotomous variables.

2. The Logistic Regression model

The Logistic Regression model is a powerful tool of data analysis which can help decision makers to identify the market segments that are more likely to respond to certain marketing actions. The logistic regression known as the regression with a binary dependent variable is used in a similar way with the Discriminant analysis, but in this case the independent variable could be also nominal ones (binary or categories). In this case the probability of an event occurring is calculated in a similar mode with multiple regressions but the values of the dependent variable in the case of Logistic Regression fall between 0 and 1. Thus the relationship between the dependent and independent variables is described by an S-shaped curve: for low levels of independent variables, the probability to occur the dependent variable is zero. As the levels of independent variables increase, the probability of the dependent variables increases on an S-curve but cannot exceed the level 1 of probability (Hair et al. 1998, 277).

Logistic regression is used in social sciences as an alternative technique to ordinary least scores (OLS) used in traditional regression models due to the fact that

often the researches regarding people behaviours use dichotomous variables instead of continuous variables. Its main advantage consists in a less restrictive application due to the absence of hypotheses regarding the multivariate normal distribution and equal variances and covariances for all variables. In this respect the logistic regression model could be used for categorical variables, which are often used in sociological surveys (Peng and So 2002, 31).

The logistic regression model is based on the mathematical concept of "logit", which is the natural logarithm of an odds ratio, where the odds is the ratio of probability of a certain event Y happening (p) to probability of Y not happening (1-p). Thus the model (see equation 1) is based mainly on dichotomous variables (Peng et al. 2002, 4).

$$Logit\,(Y) = \ln(odds) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta x \tag{1}$$

where Y is the dependent variable, x is the independent variable, α is the intercept and β is the regression coefficient.

Starting from the above relationship we can calculate the probability of an event occurring: P(Y=1/x).

$$P(Y = 1/x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \tag{2}$$

where e = the base of natural logarithms (aprox 2.72)

Extending the above model to more predictors, we can obtain the binary logistic regression model with multiple regressors (Peng and So 2002, 35):

$$P\,(Y = 1/x_1, x_2, x_3, ..., x_k) = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k}} \tag{3}$$

The predictors (x_i) could be binary or multinomial variables. In the case of multinomial variables the categories of these variables are transformed into binary variables. Thus the initial categories of multinomial variables are transformed using Dummy variables while one of these categories is considered the basis, for which all the answers are zero.

3. Applying the Binary Logistic Regression model for data analysis

Even if the application of the Binary Logistic Regression model seems to be quite complicated, the information systems for statistical data processing could help the researchers, the algorithms of data processing being very easy to use.

In the following section of this article we are going to exemplify the applying of the Binary Logistic Regression model using the SPSS (Statistical Package for Social Sciences) system.

In marketing researches it is a usual practice to find certain behaviours using multiple answers nominal scales that have a checklist of answers from which the respondent can choose one or more answer categories. Such questions are defined in data bases as dichotomous variables, which can be easily used in logistic regression models as far as other multivariate data analysis models ask for continuous variables. Thus, the binary regression models could put into relationship a future specific action with certain present behaviours. Based on such dependences the companies could establish proper marketing strategies meant to target specific market segments.

For exemplification we are going to use a research conducted among students that are accommodated in the Transilvania University's residence halls. The dependent variable (Y) is the intention of students to leave the residence hall if they will have such an opportunity and the independent variables (X) are the reasons for which they have initially chosen the residence hall. These variables have been introduced into the questionnaire by using a nominal scale with multiple choices. Both the dependent and independent variables are dichotomous variables. The answers' distribution counted by the reasons for which the students have chosen the residence hall is presented in Table 1.

		Responses		Percent of Cases
		N	Percent	
$Reasons[a]	Good prices	267	34.0%	68.8%
	Good accommodation facilities	156	19.9%	40.2%
	Pleasure to live with colleagues	160	20.4%	41.2%
	The residence halls location	78	9.9%	20.1%
	Not enough money	124	15.8%	32.0%
Total		785	100.0%	202.3%

a. Dichotomy group tabulated at value 1.

Table 1. *Reasons for choosing residence hall*

A number of 388 students answered to the above question, from which we obtained 785 answers. The reason "Good prices" counted the biggest number of answers, which represent 34% from the total number of answers. This reason has been chosen by 68.8% of respondents. The smallest number of answers was counted by the residence hall location and not enough money for other accommodation.

As regards the dependent variable we found that only 5.5% of respondents intend to choose another accommodation if they will have such an opportunity. Trying to find the influence of the above reasons on this intention we used the Binary Logistic Regression model with the help of SPSS system.

The model parameters (α and β) are presented in Table 2. We can see that for the first 4 reasons the parameters have negative values, which means a decreasing in the probability to obtain value 1 for the dependent variable, while the reason not enough money has a positive coefficient which means an increasing in the probability to change the residence hall with other accommodation. Such a result is expected if we take into consideration that the first 4 reasons are positive one while the last one is negative. Those students chose the university's residence hall because they face with a restriction: they didn't have money for other accommodation so that they are ready to leave the residence hall if they will have money.

	B	S.E.	Wald	df	Sig.	Exp(B)
Good prices (pr)	-,279	,466	,358	1	,550	,757
Good accommodation facilities (fac)	-1,397	,635	4,843	1	,028	,247
Pleasure to live with colleagues (col)	-,205	,487	,176	1	,674	,815
The residence halls location (loc)	-,072	,578	,015	1	,901	,931
Not enough money (mon)	,650	,472	1,899	1	,168	1,916
Constant	-2,464	,506	23,741	1	,000	,085

a. Variable(s) entered on step 1: v1.1, v1.2, v1.3, v1.4, v1.5.

Table 2. *Variables in the Equation*

Based on the results presented in Table 2 we can obtain the equation of our model:

$$P(Y=1/x_1,\ldots,x_5)=\frac{e^{-2.464-0.279\cdot pr-1.397\cdot fac-0.205\cdot col-0.072\cdot loc+0.65\cdot mon}}{1+e^{-2.464-0.279\cdot pr-1.397\cdot fac-0.205\cdot col-0.072\cdot loc+0.65\cdot mon}} \quad (4)$$

In table 2 are also presented important information regarding the real influence of the independent variables on the increasing or decreasing of probability to occur the event measured by the dependent variable (see column Exp (B)). The Exp (B) values indicate how many times increases the probability of the event to occur when the predictor increases its value with one unit. By subtracting 1 from this value (Exp (B)-1) we obtain the changing ratio in the probability, which can be positive or negative. This ratio could be transformed in percentages for a better explanation of these influences.

For example the ratio of probability changes of the dependent variable in the case of a student that chose the residence hall for good prices (pr) is -0.243 (0.757-1). This means that for such a student the probability to leave the residence hall decrease with 24.3%. Following the same algorithm for the ones that chose this place for good accommodation facilities the probability to change the accommodation decrease with 75.3% (0.247-1= - 0.753). For the students that didn't have enough money for other accommodation the probability to leave the residence hall increase with 91.6% (1.916-1=0.916).

In table 2 there are also presented other useful information regarding the individual signification of the model's parameters. Thus in order to be significant a parameter have to obtain a very low significance level (Sig. < 0.05) and the Wald value to be bigger than 0.5. In our case we can see that only the variable "Good accommodation facilities" recorded a significance level smaller than 0.05. In conclusion the influence of the other variables could not be considered statistically significant and these ones have to be excluded from the model. In this respect the Forward Stepwise Regression method should be selected in SPSS and the system will select only the significant regressors.

The significance of the entire model could be appreciated looking to the R square value, which indicates the ratio of the regressors influence in the total variance of the dependent variable. This value has to be closed to 1. For our example the R square calculated with different methods (Cox & Snell or Nagelkerke R Square) has small values, which means that the model has a low explanation power and other variables should be considered (see Table 3).

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	158,559[a]	,028	,079

Table 3. *Model Summary*

In conclusion, the model has helped us to identify some patterns of the students' behaviours and the probability of these ones to influence future decisions.

But the results have to be carefully interpreted when the decisions is made because a poor statistical significance of these results could lead to wrong decisions. In this respect from scientifically point of view a model cannot be considered if it is not statistical significant. Therefore a stepwise method for variables selection is recommended and the majority of information systems could perform such processes.

4. Conclusions

Taking into account the results obtained by applying the Binary Logistic Regression model on the answers given by students we can conclude that the university have to concentrate its marketing efforts on improving the quality of accommodation services and on creating a friendly climate in the residence halls in order to obtain a high level of students' satisfaction. A simple concentration on low prices and those students that cannot afford another accommodation is not a proper strategy from the marketing point of view. Once this improvements will be put in practice a strong communication strategy should performed in order to inform the students and future candidates about the new accommodation facilities and conditions. The communication strategy should take into consideration those channels used by students including the social media, which have become very popular especially among young people (Chitu & Tecau 2012, 35).

In spite of these results, the decisions have to be taken with prudence because the results have a poor statistical significance. For this reason the model have to be reconsidered in order to find other variables that have a significant influence on the dependent variable and contribute to a higher explanation of this variable's variance.

Beyond the limits of the above example it has been used to underline the possibility to obtain information that is not fully relevant because it has a poor statistical significance. For this reason, in such situations data should be reconsidered by finding new variables that can improve the quality of the model and its statistical significance. When these conditions are met the applying of the Binary Logistic Regression model in marketing researches could create a strong support for decision makers. They can quantify the probability of an event occurrence based on independent variables that express in a binary way certain behaviours or socio-demographic characteristics of the customers. If these variables have a statistically significant influence on the dependent variable the decision makers could concentrate their marketing efforts on those marketing segments that are the most susceptible to react positively to certain marketing actions. The multivariate analysis could be enhanced by using bivariate methods like crosstabulation in order to find patterns when the dependent variable is put in relationship with every independent variable considered one-by-one.

In conclusion, this instrumental research could be useful both for marketing practitioners, academicians and researchers because it explains in details how to apply the Binary Logistic Regression model and how to interpret the results. It is also stressed the necessity to avoid certain mistakes regarding the results interpretation when these ones have a poor statistical significance. Also the doctoral students could benefit from this research in their efforts to explain the results of their own applied researches.

5. References

Chitu, Ioana B. and Alina S. Tecau. 2012. "Issues related to social network advertising." *Bulletin of the Transilvania University of Braşov Series V: Economic Sciences,* 5 (54), No. 1: 31-36.

Florescu, Constantin, Petre Mâlcomete and Nicolae Al. Pop (coord.). 2003. *Marketing. Dictionar explicativ (Marketing. Explanatory dictionary).* Bucharest: Economic Publishing House.

Hair, Joseph R. jr., Rolph E. Anderson, Ronald L. Tatham, and William C. Black. 1998. *Multivariate Data Analysis,* 5th edition. New Jersey: Prentice Hall International, Upper Saddle River.

Ispas, Ana, Denisa G. Boboc, Andreea G. Marica, George B. Nastasoiu and Cristina V. Vijoli. 2015. "City Brand – Identity, Positioning and Image. Case Study: City of Brasov." *Romanian Journal of Marketing,* 1/2015: 22-31.

Kotler, Philip and Kevin L. Keller. 2006. *Marketing Management,* 12th edition. New Jersey: Pearson Education Prentice Hall, Upper Saddle River.

Lefter, Constantin. (coord.) 2006. *Marketing.* Vol. II. Brasov: Transilvania University of Brasov Publishing House.

Peng, Chao–Ying Joanne and Tack-Shing Harry So. 2002. "Logistic Regression Analysis and Reporting: A Primer". *Understanding Statistics,* I (1): 31-70.

Peng, Chao–Ying Joanne, Kuk Lida Lee and Garry M. Ingersoll. 2002. "An introduction to Logistic Regression Analysis and Reporting". *The journal of educational research,* 96 (1): 3-14.

Todor, R.D. 2014. "The importance of branding and rebranding for strategic marketing". *Bulletin of the Transilvania University of Braşov Series V: Economic Sciences,* 7 (56) No. 2: 59-64.