A COMPARISON BETWEEN MULTIVARIATE AND BIVARIATE ANALYSIS USED IN MARKETING RESEARCH

Cristinel CONSTANTIN[1]

Abstract: *This paper is about an instrumental research conducted in order to compare the information given by two multivariate data analysis in comparison with the usual bivariate analysis. The outcomes of the research reveal that sometimes the multivariate methods use more information from a certain variable, but sometimes they use only a part of the information considered the most important for certain associations. For this reason, a researcher should use both categories of data analysis in order to obtain entirely useful information.*

1. Introduction

The quality and the quantity of the information obtained from marketing research are very important for the decision makers of a company. For this reason, the data collected are analysed with various methods meant to obtain the information needed. But sometimes the researchers prefer to use certain methods according to their knowledge or try to overrate the importance of multivariate methods. Our research aim is to compare some multivariate and bivariate methods starting from the hypothesis that these methods should be complementary in data analysis.

2. Literature review

Multivariate analysis is considered as all statistical methods that simultaneously analyse multiple measurements on each individual or object under investigation. It deals with multiple combinations of variables, which are put into practice by using various multivariable methods [1]. These variables may be correlated with each other, and their statistical dependence is often taken into account when analysing such data. Response variables are often described as random variables, being often described by their joint probability distribution [2].

The impressive development of information technology allowed the scientists' access to multivariate analysis, which needs a large amount of data processing and very complex algorithms. But using multivariate analysis has become a strong need for decision makers taking into consideration the complexity of markets and consumer behaviours. Some authors consider that any problem that is not analysed on a multivariate basis is treated superficially and the multivariate

---

[1] Dept. of Economic Sciences and Business Administration, *Transilvania* University of Braşov.

analysis will be predominant in the future. It will change the way in which researchers think about problems and how they design their research [3].

Usually, the multivariate analysis uses a linear combination that seems to be too simple at a first glance but this one has at least to major advantages: it has a mathematical applicability and often performs well in practice. The statistical prerequisites of these methods are basic familiarity with the normal distribution, t-tests, confidence intervals, multiple regression, and analysis of variance [4].

Multivariate analysis techniques can be classified into two major categories: dependency techniques and interdependency techniques. The former consist of techniques in which a variable identified as dependent variable depends on the variation of other variables, identified as independent variables. On the other hand, interdependency techniques involve the simultaneous analysis of all the considered variables. [5]

In literature, multivariate methods are often considered as extensions of univariate or bivariate analysis, but their value can be higher as long as many variables are put together into analysis [6]. However, both multivariate, bivariate and univariate analyses are very useful in order to obtain the research outcomes needed by decision makers.

Starting from these considerations, for a researcher it is very useful to know the utility of every analysis technique for the research analysis, but also their shortcomings regarding the results' interpretation.

3. Research methodology

In order to emphasize several characteristics of some data analysis methods, we performed an instrumental research. This kind of research is used for the testing and validation of the methods and instruments used in marketing research [7]. In this respect, the main objective of our research was to make a comparative

analysis of certain bivariate and multivariate data processing methods. The results were compared in order to find and underline some possible shortcomings or strong points of every method.

The research was made using a database resulted from a marketing research among car owners from Brasov-Romania, whose main aim was to identify customer satisfaction regarding different car brands. The sample counts 100 car owners from Brasov, over 18 years old, randomly selected. The data were processed using the SPSS system, using both bivariate and multivariate methods.

4. Discriminant analysis

Discriminant analysis is used to determine which continuous variables discriminate between two or more naturally occurring groups. It is also used to determine which variables are the best predictors of including an individual into a certain group resulted from the dependent variable. Discriminant analysis answers the question: can a combination of variables be used to predict group membership? [8]. In the set of variables used, one is considered as dependent (being nominally scaled) and is put in relationship with several independent variables that are interval or ratio scaled [9]. This method is frequently used to classify the respondents in two groups, according to a dichotomous dependent variable. In such a case, a discriminant function is computed, which helps us predict the proper group of an individual according to his characteristics, given by the independent variables [10].

In order to perform the discriminant analysis, we used the following variables: the respondents' satisfaction regarding their Dacia car as dependent variable; the attitudes about the safety and the price-quality ratio of the automobile and the maintenance costs in the last year as independent variables. The dependent variable is nominally measured, using a dichotomous scale with "Yes" and "No" answers. The independent variables

are measured with the ratio scale (maintenance costs) and numeric interval scales with 5 levels (price-quality ratio and safety). Table 1 presents the descriptive statistics of the two groups (satisfied and unsatisfied people with the quality of Dacia cars).

Group Statistics Table 1

Are you satisfied with the quality of the car that you have?		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
No	Price-quality ratio	3.38	1.188	8	8
	Safety	2.38	1.061	8	8
	Maintenance cost	1200	968.061	8	8
Yes	Price-quality ratio	4.37	.926	27	27
	Safety	4.04	.980	27	27
	Maintenance cost	734.07	595.013	27	27
Total	Price-quality ratio	4.14	1.061	35	35
	Safety	3.66	1.211	35	35
	Maintenance cost	840.57	709.283	35	35

The answers were collected from 35 people that are Dacia owners. The means recorded by the two groups for the independent variables are quite different. Those who are satisfied with the quality of their vehicle have the average satisfaction regarding the quality-price of the car of 4.37 points on a 5 level scale (5-very satisfied). In the other case (those who are not satisfied with the quality of the vehicle), the mean is 3.38 points on the same scale. The average degree of satisfaction regarding the car's safety is also higher for those who are satisfied with the car quality than those who are not satisfied. There is also an inverse correlation between the general satisfaction and the maintenance costs, the satisfied respondents experiencing lower maintenance costs [11].

The test of statistical significance of the above differences shows that the satisfaction with the price-quality ratio and satisfaction with the car's safety have a significant influence on the general satisfaction according to the results of the Fisher test (see table 2).

Tests of Equality of Group Means Table 2

	Wilks' Lambda	F	df1	df2	Sig.
Price-quality ratio	.840	6.272	1	33	.017
Safety	.658	17.132	1	33	.000
Maintenance cost	.922	2.804	1	33	.103

Thus, both variables have a significant discriminant power as the calculated "F" values are higher than the theoretical values obtained from the Fisher distribution table (sig.$\leq$ 0.05). The other variable, the maintenance costs, does not significantly influence the general satisfaction with Dacia cars (sig.> 0.05). The test results show that "safety" has a higher discriminant power because its Wilks' Lambda coefficient is lower and the F value is higher than in the case of "price-quality" variable.

The above results are also reflected in the equation of discrimination function. In this

function, the coefficient of the variable "maintenance costs" has zero value (table 3)

Canonical Discriminant Function Coefficients   Table 3

	Function
Price-quality ratio	-.149
Safety	1.002
Maintenance cost	.000
(Constant)	-2.656

According to the results presented in table 3, the equation of the discriminant function is:

$$y = -2.656 - 0.149 \cdot x_1 + 1.002 \cdot x_2$$

(1)

Where, x_1 is the satisfaction level regarding the price-quality ratio and x_2 is the satisfaction level regarding the car's safety.

Applying this function on the existing respondents, we obtained the results presented in table 4.

Classification Results   Table 4

		Are you satisfied with the quality of the car that you have?	Predicted Group Membership		
			No	Yes	Total
Original	Count	No	4	4	8
		Yes	2	25	27
	%	No	50	50	100
		Yes	7.4	92.6	100

Out of eight respondents who said they are dissatisfied with the quality of the car, four respondents (50%) should be part of the satisfied group according to the discriminant score. Thus only 50% of respondents are correctly classified. In the other case (people satisfied with their cars' quality), 92.6% of respondents are correctly classified according to the discrimination function. In this respect, we can conclude that the obtained function could give quite good results in identifying satisfied customers according to their attitudes regarding the quality price ratio and the car's safety.

An analysis similar to the above one could be made using bivariate methods, such as the cross tabulation and specific statistical tests (see table 5).

The relationship between the general satisfaction and the attitudes towards safety   Table 5

Are you satisfied with the quality of the car that you have?	How satisfied are with the safety of your Dacia car?					Total
	1-Very unsatisfied	2	3	4	5-Very satisfied	
NO	100,0%	50,0%	36,4%	11,1%		24,3%
YES		50,0%	63,6%	88,9%	100,0%	75,7%
Total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

In table 5 we can see that the majority of people satisfied with the safety of their automobile are generally satisfied with the quality of this car. On the other hand, the respondents who are not satisfied regarding the car's safety are generally not satisfied with the quality of their car.

The same situation could be found in the case of the attitudes regarding the price-quality ratio, but the relationship seems to be not as strong as in the above case (see table 6). We can see that there is a high percentage of respondents who are satisfied with the price-quality ratio but are not generally satisfied with the quality of their cars.

The relationship between the general satisfaction with Dacia cars and the attitudes towards their piece-quality ratio   Table 6

Are you satisfied with the quality of the car that you have?	How satisfied are you with the price-quality ratio of your Dacia car?				Total
	2	3	4	5-Very satisfied	
NO	50,0%	50,0%	30,0%	11,1%	26,3%
YES	50,0%	50,0%	70,0%	88,9%	73,7%
Total	100,0%	100,0%	100,0%	100,0%	100,0%

The difference between the two situations above is more evident when we apply some statistical tests. Therefore, by applying both Chi-square test and Kolmogorov-Smirnov test, which are suitable in these cases, we obtained a significant relationship between safety and the general satisfaction with Dacia cars, but this relationship is not significant in the case of the price-quality ratio.

As regards the third variable, ("maintenance costs"), which is measured with a ratio scale, the difference between the satisfied respondents and the unsatisfied ones cannot be considered statistically significant (sig.>0.05).

Comparing the discriminant analysis as a multivariate method with the bivariate analysis performed above, we can conclude that the results are similar, but they are different in some details. For example, whereas in the bivariate analysis the satisfaction regarding the price-quality ratio is not useful for explaining the general satisfaction regarding the analysed car, the multivariate analysis uses the information given by this variable to classify respondents according to their general satisfaction. Moreover, the multivariate analysis allows us to make predictions about a certain individual according to his or her values regarding the independent variables by using the discriminant function. This kind of predictions is not available in the case of bivariate analysis, which is a descriptive method.

5. Homogeneity analysis

Some of the multivariate analyses have the disadvantage of using ratio or interval variables, but in most of the situations the questionnaires contain a lot of nominal variables. For these kind of variables, including the characterization ones (gender, age, incomes etc.), some reduction methods are used, one of the most popular being the Homogeneity analysis. Also known as the Multiple Correspondence Analysis, this method makes complicated multivariate data accessible by displaying their main regularities in pictures such as scatter plots [12]. It provides an easily interpreted perceptual map that jointly shows the relationship between the categorical variables, which is not available through the

traditional method of using Chi-Squares on a bivariate level of analysis [13]. What the technique accomplishes is to scale the N objects (map them into a low dimensional Euclidean space) in such a way that objects with similar characteristics are close

together, while objects with different characteristics are relatively far apart [14].

In our research regarding automobiles, we used the Homogenity analysis in order to emphasize the relationships between gender, age, income and the automotive brand respondents owned.
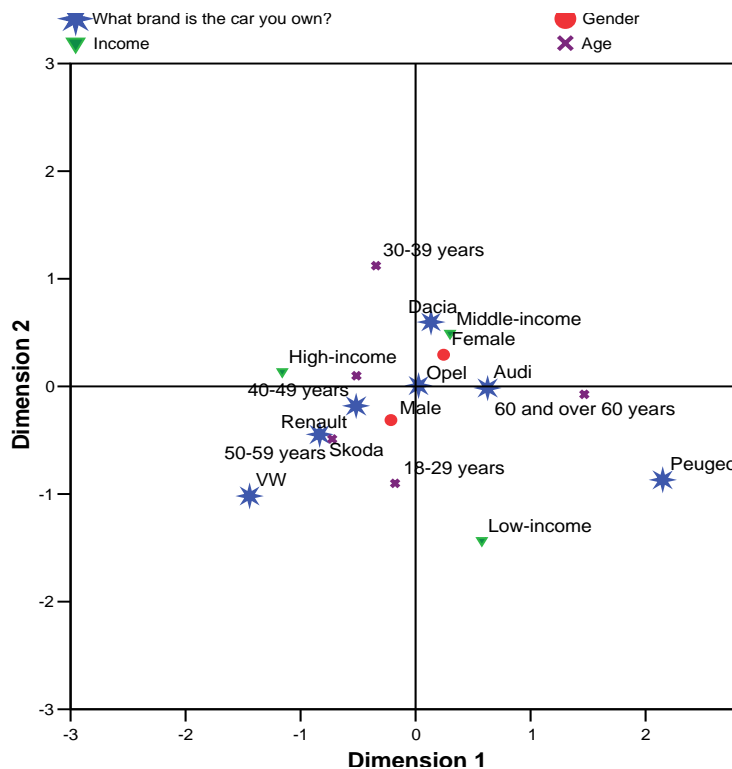
Fig. 1. *Homogeneity analysis between gender, age, income and car brand*

In Figure 1, the categories of analyzed variables were represented by dots. The dots' emplacement shows that the brands Renault, Skoda and VW are owned mainly by males, people with high incomes (over 2500 RON) and aged between 18 and 29 years, and between 40 and 60 years old. Brands like Dacia, Opel and Audi are owned mainly by women, people with medium income (between 1200 and 2500 RON) and aged between 30 and 39 years. The last brand (Audi) is an anomaly in this correlation between variables because the respondents' income does not match the high price of these automobile, and also

the high maintenance costs. This one could be explained by the small number of respondents who indicated this brand, which cannot be considered statistically significant [15].

Some categories are quite isolated on the chart, being weakly correlated with the other categories. These ones are Peugeot cars, which recorded a small number of answers and also other categories, such as the respondents aged over 60 years and the ones with low income (below 1200 RON).

Table 7 shows the bivariate analysis, in which the car brand was cross tabulated with every characterization variable

(gender, age and income).

We can see that the same associations from the above analysis are quite hard to interpret as the relationships are computed in different tables.

If we look at the association between gender and car brand, we can see that females usually drive Dacia, Opel and Renault brands, while males drive Dacia, Opel, Skoda and Volkswagen. The results are quite different from the ones revealed by the Homogeneity analysis. The last one shows a quite long distance between males and Dacia cars, even if the percent of males that own this brand is the highest one. This difference could be explained as the Homogeneity analysis takes into consideration the dominant characteristic of every group; in this case, the females' percentage is significantly higher than the males' one. In conclusion, even if both males and females own the Dacia brand to a large extent, this phenomenon is more frequent in the case of females than males.

Crosstabs between car brand and
the main respondents' characteristics   Table 7

	Dacia	Skoda	Opel	VW	Peugeot	Audi	Renault	Other	Total
Gender									
Female	44,9%	6,1%	10,2%	6,1%	4,1%	6,1%	8,2%	14,3%	100,0%
Male	31,4%	13,7%	17,6%	9,8%	5,9%	5,9%	3,9%	11,8%	100,0%
Age (years)									
18-29	35,0%	10,0%	5,0%	10,0%	-	5,0%	10,0%	25,0%	100,0%
30-39	40,9%	9,1%	22,7%	-	-	9,1%	9,1%	9,1%	100,0%
40-49	33,3%	16,7%	11,1%	11,1%	5,6%	5,6%	11,1%	5,6%	100,0%
50-59	41,2%	11,8%	11,8%	23,5%	-	-	-	11,8%	100,0%
over 59	39,1%	4,3%	17,4%	-	17,4%	8,7%	-	13,0%	100,0%
Income									
Low	18,2%	9,1%	18,2%	4,5%	9,1%	9,1%	9,1%	22,7%	100,0%
Middle	48,1%	7,4%	11,1%	5,6%	5,6%	5,6%	3,7%	13,0%	100,0%
High	33,3%	16,7%	16,7%	16,7%	-	4,2%	8,3%	4,2%	100,0%

From another point of view, the Homogeneity analysis takes into consideration the simultaneous relationships between variables, so that only the most important information is extracted from every variable. It means that some information is lost and the results should be interpreted cautiously. For this reason it is recommended to perform both bivariate and multivariate analysis and to extract the essentials from every outcome. For example, it is a mistake to say that males or people in the 50-59 years age range do not own Dacia cars as could be interpreted from Figure 1. According to the bivariate analysis, we can conclude that most of the respondents, irrespective of their characteristics own a Dacia car. But as differentiation characteristics, these respondents are closer to Skoda and Opel than other categories (e.g. females and people over 60 years old).

6. Conclusions

Looking at the results that can be obtained using the multivariate and bivariate analysis, we can conclude that both categories have strong points but also shortcomings. Usually, the multivariate methods give us a synthetic image of the relationships among more than two variables, helping researchers obtain certain information very easy. Sometimes, as we found in the case of the Discriminant analysis, a multivariate method can use the information from a variable even if this

information cannot be considered statistically significant using a bivariate test like chi-square or Kolmogorov – Smirnov.

Using other methods, like the Homogeneity Analysis, it is better to complete the analysis with bivariate methods in order to find some losses of information.

For researchers and decision makers, it is important to use the multivariate and bivariate methods as complementary, as these ones are not interchangeable in all situations. These results confirm our starting hypothesis.

Finally, the researchers should improve the communication with the research beneficiary in order to understand their needs and to discuss the final report with the same people in order to clarify the outcomes and to avoid possible misunderstandings [16].

Acknowledgements

References

1. Chandra, S., Menezes, D.: *Applications of Multivariate Analysis in International Tourism Research: The Marketing Strategy Perspective of NTOs*. In Journal of Economic and Social Research 3(1), 2001, pp.77-98. See [5]

2. Constantin, C.: *Sisteme informatice de marketing (Marketing Information Systems)*. Braşov, Infomarket, 2006. See [10].

3. Cooper, D., Schindler, P.: *Business Research Methods*. McGraw Hill, International edition, 2006. See [9]

4. de Leeuw J., Mair, P.: *Gifi Methods for Optimal Scaling in R: The Package homals*. In Journal of Statistical Software 31, (4), 2009, pp. 1-21. See [14]

5. Ferencová, M.: *Verbal communication in corporation and means of consigning the knowledge*. In: Management 2008 (Part II.) (Eds.) Róbert Štefko, Miroslav Frankovský University of Prešov in Prešov, 2008. - ISBN 978-80-8068-849-3 pp. 30-38. See [16].

6. Hair, J. jr., Anderson, R. et al.: *Multivariate Data Analysis*. 5th edition, Prentice Hall International, 1998. See [1], [3], [6].

7. Lefter, C.: *Cercetarea de marketing (Marketing research)*. Braşov, Infomarket, 2004. See [7]

8. Michailidis, G., de Leeuw, J.: *The Gifi System of Descriptive Multivariate Analysis*. In Statistical Science 13, (4), 1998, pp. 307–336. See [12].

9. Pirvu, V., Constantin, C.: *A multivariate analysis of the Romanian car holders' attitudes and behaviours*, In ASPECKT Journal, 6, 2010, pp. 63-70. See [11], [15].

10. Poulsen, J., French A.: *Discriminant function analysis*. Sage Publication, 2001. See [8]

11. Rencher, A.: *Methods of multivariate analysis*. 2nd edition, John Wiley & Sons, 2002. See [4].

12. Schimmel, K., Nicholls, J.: *Segmentation Based On Media Consumption: A Better Way To Plan Integrated Marketing Communications Media*. In Journal of Applied Business Research 21 (2), 2005. pp. 23-36. See [13]

13. Stevens, J. P.: *Applied multivariate statistics for social sciences*. 5th edition, Lawrence Erlbaum, 2001. See [2]