Analysis of seasonal discounts in small markets using statistical methods

ELMIRA KUSHTA[1], MIFTAR RAMOSAÇO[2]
[1]Department of Mathematics, Faculty of Technical Sciences, University "Ismail Qemali", Vlore, ALBANIA
[2]Department of Mathematics, Faculty of Technical Sciences, University "Isamil Qemali", Vlore, ALBANIA

Abstract: In this paper we consider the apparent changes in consumer spending in a real market (Vlora County). By calculating the latent variable of discounts as it is done, we do not interpret the net of expenses. We found that discounts stimulate spending behavior, which shows that the consumer buys the minimum as well as the highest prices and not with the right budget. Essentially, the ecological steps here have functioned as data-oriented modeling that reduces subjective or am approaches to econometric and marketing analysis. The complexity of separate actions and then its dimensions and reduced the model of conclusion and improved analysis.

Keywords: Probit model, consumer behavior, distributions, logistic function.

Received: September 28, 2022. Revised: August 30, 2023. Accepted: September 21, 2023. Published: November 2, 2023.

1. Introduction

Consumer behavior is an opinion formation process which is difficult to be studied quantitatively and rather complex as evidenced in [1].
The study of concrete economic environment unavoidably face challenges to the scholars. Standard questionnaires that aims gathering information from social or economic mediums include different types of variables, non-numerical responses, questionable answers, missing or incomplete records etc. Usually the inquiries might be organized and held in different moment of times. Finally they must be included in modeling say linear multivariate functions

In a formal approach they are assumed to act rationality in their decision making by optimizing some utility function, but this last cannot be measured directly. Meanwhile the assumption of rationality does not hold always which is clarified by behavioral theories as discussed in [2], [3] etc. Behavior seems to be too complex to be studied and analyzed by deterministic methods. Even so, researchers and scholars have outdone this difficulty by using statistical tools and probabilities facilitating modeling in this case as presented in many textbooks. The second problem is related to the tangible set of the variables included in the models. Again, standard models belong to the standard systems and in real ones there is a considerable difference Mixed calculation using econometric optimization and network dynamics have been developed as for example in [4] and many other applications. Generally, the consumer's decision making process in buying is complex but econometrically known and measurable. But by carefully using simple analytic tools it is possible to avoid the complexity of the model, to control extra errors added during calculation phase and improve overall calculation. In our recent research in the analysis of consumer behavior in district of Vlora, we considered such specifics as an important step [1], [2] etc. Aside of general models and regressions, practical calculation have demonstrated their capacities to describe consumer behavior in specific systems as in [5], [6] and many others. In our recent work [7] we applied a logistic regression to identify the consumer profile in a specific area, the factors affecting their behavior and other parameters characterizing the system of consumer attitudes and activities. Therein we have focused our attention in the fat that the stationary of the state should be considered in the framework of advanced analysis elsewhere proposed in [8] and [9] with mathematical reviews in [10]. In reference [11] a more advanced technique have been reported dealing with complexity in the behavioral models. From those and other references that we are not listing here, we acknowledged the importance of folded econometric and mathematical details for quantitative consideration for such systems. Detailed analysis on those aspects are provided in many articles-guides and statistical books as [3] or [4]. In this case the problems could be overcomes if we adjusted correctly the sample size or adopt a suitable sampling method. In the case where the above step is not suitable or even impossible, the factorial and descriptive analyses could be used as recommended in standard procedures to manage the sampling error, [5] and general consideration [13]
This study is intended to evaluate a marketing aspect as discounts for example, by specifically considering the nature of the state of the system, the possible presence of not-apparent factors as latent effects or hidden variables etc.

2. Optimization by maximum likelihood.

Referring to the linear discriminate treatment above, we have evidenced that each of the posed binary variables is suitable to identify the probability of customer behavior [12].The estimate here is a benchmark estimate that should be pushed further with the likelihood function optimization technique through this procedure on the candidates proposed by matching OLS or W-OLS [7]. In most cases, we have applied specially built routines, so the calculation part is chronologically recorded here. Of course, calculations based on sequential forms are the alternative behavior since most software have built-in functions. [20] Assessing that the deviations are for accidental reasons.

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (1)$$

The likelihood estimator has the form

$$ln = log_e g(X_1, \dots, X_n)$$

$$\ln\left(\left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{\sum w_i}{2}} e^{\frac{\sum w_i (x-\mu)^2}{2\sigma^2}}\right) = -\ln[\sigma^2 2\pi - \frac{1}{2\sigma^2}\sum(x-\mu)^2 w_i] \qquad (2)$$

$$P(Y_i = 1) = \pi_i = \frac{exp(\beta_0+\beta_1 x_i)}{1+exp(\beta_0+\beta_1 x_i)} \qquad (3)$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_{p-1} \end{bmatrix}_{p\times 1} \quad X = \begin{bmatrix} 1 \\ X_1 \\ \dots \\ X_{P-1} \end{bmatrix}_{P\times 1} \quad X_i = \begin{bmatrix} 1 \\ x_{i1} \\ \dots \\ x_{i,p-1} \end{bmatrix}_{p\times 1} \qquad (4)$$

$$\beta = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1}$$
$$X_i{}'\beta = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1}$$

$$E\{Y_i\} = \pi_i = \frac{exp(X_i{}'\beta)}{1+exp(X_i{}'\beta)} \qquad (5)$$

Again the best regression (optimal decomposition coefficients) is found if their infinitesimal change does not change the cost function.

$$p_j = \begin{cases} \dfrac{e^{X\beta_j}}{1+\sum_{j=2}^{m} \exp(X\beta_j)} & j = 2\dots m \\ \dfrac{1}{1+\sum_{j=2}^{m} \exp(X\beta_j)} & j = 2\dots m \end{cases} \qquad (6)$$

We have tried to follow the answers in the form "expenses of category x are in group I " where we have divided the expenditure classes into subgroups. Accordingly, we obtained m=3 by grouping the expenditure variants in 3 classes. In this attempt, we have not found a better answer. [18]The following two-valued case considers the classical case with binary responses.

3. Identification of the local optimal model

The treatment here is local in the sense that the treatment is done unilaterally. [23]We have fixed a variable or several of them as responses, and we try to find the best model i.e. the causal variables that lead to those responses. In the analysis of the importance of the model and the results obtained, we find that there is evidence of causality in the decision-making to spend more on basic goods than on all others.

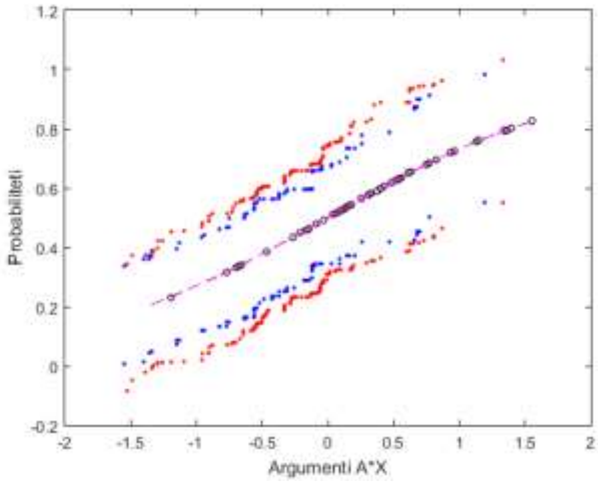

Figure 1: Logistic regression for different sets of variables.

Black circles log regression for variables X1:X6 with pink dots, regression for X1:X3. The blue and red dots give the 95% confidence interval.
In figure 1 we have presented the result as well as the margin of error of the algorithm. [15]We note that in this approximation:

- The probability of the outcome (outcome) extends well in the space of allowed values [0,1]
- There is a defect at the edges as it does not respond to a complete sigmoid but only the central part.

This behavior indicates that the domain of determination is incomplete at the edges and in this case a good strategy is seen to include a continuous variable in the predictors X since they are all categorical. Likewise, the removal of some variables causes us to change this level of inaccuracy by suggesting the strategy of tests for each one, we have concluded linearly as can be seen in table 1, where the values of Eigen-significance p, error (deviation) and statistics of factual assessments are presented. It is found that these parameters vary from variable to variable and have significant values that question the completeness of the model.

Table 1: Variables of the model

	Gender	categorical	1,2	Possibly differenced effects
	Age	categorical	1,2,3,4	Possibly differenced effects
before discounts announcements	Average visit on the market	categorical	[1-30]	informative variable
	Total visits	nominal	[1-30]	informative variable
	Average expenses	nominal	Real Value	Depended variable
	Total expenses	nominal	Real Value	Depended variable
	Having Market Cards	binary	[1,0]	increase chance of contacts
	Having contacted by phone	binary	[1,0]	increase chance of contacts
after discount announcements	Total Visits	nominal	[1-30]	Informative variable
	Average visits	nominal	[1-30]	Informative variable
	Average expenses	nominal	Real Value	Depended variable
	Total expenses	nominal	Real Value	Depended variable

The state of the system is characterized by average purchasing expenses, number of visits in the market. [23]Therefore the aim of the analysis is focused in the evolution of those quantities if sales or discounts are applied. [7] We realized it by analyzing the distribution of the response quantity measured in those two states respectively, and applying a probity model to estimate the utility underlining the dynamics observed.

4. Obtaining Meaningfulness of Variable Type and Units.

It resulted that for all response variablesrelative counterparts were characterized by smooth distribution, Table 2. For predictor variables we adopted a descriptive approach by classifying direct values in categories.

Table2: Variables of the model

Predictor Variables		Value	
Variable	Type	set I	Set II
Family type	categorical	1-5	1-5
Education level	categorical	1-5	1-6
Age	categorical	1-6	1-4
Employment Status	categorical	1-3	1-2
Income Type	categorical	1-3	1-9
Gender	categorical	1-2	1-2
Total Consummator Budged	numerical	Real Value	Real Value
Initiallyvariable. Value Real/Proportion. Categorical value 1-5	Representativevariables. Real/Proportional	ProposedVariable. Real/Pro	categorical Value :1/0
Expenses for:		Common expenditure	
alimentary goods	Basic expenditure		Is dominant: 1/0
clothes			
subsistence			
Alcoholicdrinks and cigarettes	Extra expenditures		
health*			
Transport			
Communication (mob Phone calls)			
Culture and safety expenses	Qualitative Life Expenditure	Quality life and luxuryex_penditures	Is dominant
Education			
auxiliary services			
Family expenses			
luxury goods	Luxury_Expenditure		
Restaurant_Expenses			

We asked for variables to be appropriate for modeling in logistic, MIMIC and other form if they were found in a stationary state. We managed the measurement realized in the sample where an individual appears as a list of records of different type and different meaning. [19]To include all of them in an deterministic model we must unify their measurement method. Hence categorical variables were transformed using z-score method in continuous variables. $x \to \frac{x-\langle x\rangle}{\sigma(x)}$. In another step we produced new variable binary by using levels of expenses.

$$ExpencesLevel \leftrightarrow R_i \equiv \frac{Expence\_i}{Total\_Expences} : Y(R_i) = \begin{cases} 1, R_i > 0.5 \\ 0, R_i < 0.5 \end{cases}$$

This last is suitable for logistic and probit modeling

5. Minimal modeling testing

In constructing models we analyses all minimal model of the form logit(x)=a+bx and keeps for further consideration variables that showed good statistics of the minimal fit. In the table below are displayed.

Findings

By using such steps we concluded that the profile of the costumer for the economic medium analyses is

$$Costumer\ \Pr ofile = \begin{pmatrix} FamilyType \\ EducationLevel \\ Emplymnet \\ IncomeType \\ AgeGroupHoseholder \\ GenderHoseHolderer \\ BudgetLevel \end{pmatrix}$$

The most appropriate reference for variables measured or their natural units are as follows.

Cause Variables: categorical Likert type 3 or 5 values

Response variables: Binary, measuring the prevalence of an expense type

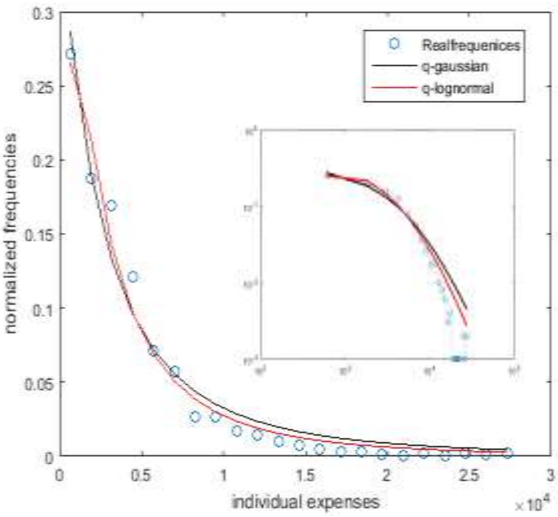

Figure 2: Distribution of visits average expenses. Normal trading period. Small picture shows log-log representation to better picture of the fit.

After realizing the fit to some expected common distributions, we observe that the parametric q-distribution had best statistics of fitting. The fitted curves are mostly q-lognormal within alpha=0.05 restriction, whereas q-Gaussian has lower statistics, but is much less sensitive to the binning assize. Accepting functions of the type (2) as best fitted distributions, one can admit that the processes underlying the expenditures dynamics are q-multiplicative, hence very complicated. From the fitted q-lognormal we obtained the parameter q=1.0001 which report a nearly stationary lognormal if multiplicative processes are determinant. In particular it does not give the opportunity to measure the level of non-stationary as the difference q-1=0.0001 is too small. But in first equation of (3) we see that q-addition involve additive and multiplicative property, so for mixed processes it seems to be more significant. For this reason we prefer q-Gaussian for analysis of such behavior. Parameters q and adjusted R-Squared are [1.6531 0.9661] for q-Gaussian and [1.0001 0.9742] for the q-lognormal fitted. Therefore q-Gaussian tells that q~5/3 that is in the boundary of definition for variances

$$\sigma_q = \frac{1}{(5-3q)\beta} \tag{7}$$

Next we considered the data for market visits and average expenditures after sales were applied. We obtain that the expenditure's distribution were found in a more stable states. The statistics for q-distributions fitted to the frequencies of consumer visits at the market again support the q-lognormal as best fitted function, but again by changing bin size we observe that q parameter in q-Gaussian changed only slowly whereas for q-lognormal it jumps from the value 1 with high margin. Therefore we consider q-Gaussians for further analysis. Q-Parameters estimated and R2 for this case are found [1.6525 0.9778] for q-Gaussian and [1.0000 0.9974] for q-lognormal. We see that the stationary parameter q is nearly the same for the two series (before and after sales) but as we explained above the observation time for the second is much lower. So we accept that the state after discounts is more stable .

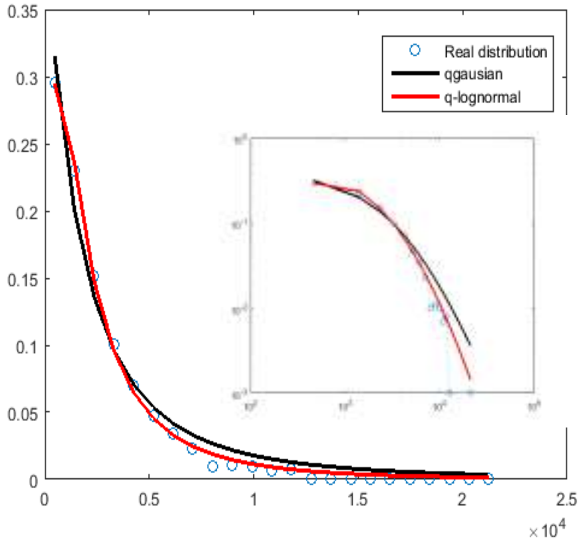

Figure 3: Distribution of average expenses (per visit) after trade discount announced.

We observe that announcing price discounts cause the number of visits to be higher and more averaged, that impose a more relaxed state which in turn could be related to a more stationary behavior or more common, typical consumer activities. In this sense we can this state for statistical analysis of the market, estimation of the representative values for expenditures and forecasting. We can simulate scenarios e.g., arrival of individuals with given set of properties using the distribution found here as PDF of the values for quantities discussed here. Moreover, we can approach this behavior as a response to a natural utility optimization and dominant factors to be typical ones, hence the disturbing terms could be neglected. Remembering that the time of post discount study has been as few as one third of the prior discount spanning period, the relaxation effect could be even more important and hence we can expect an intensive change of the probability of choice induced by discount announcement. This preliminary finding will support a more intensive change in cumulative probability CDF and therefore, the probit model could be more ascribing. In statistical aspect the result as of herein stated that the overall state is more stationary after applying discounts and we conclude that the first direct result of such marketing options in the relaxation of the system itself, and brining the distribution in the zone where the mean and variance are measurable and well defined. In this sense, marketing studies and other analysis on the consumer behavior are likely to be more realistic in the period of sales. It is very important result for the market studies in the case of small market area as local districts and limited capacities for large inquiry that impose instability of the state of the system.

The coefficients have been confirmed as different from zero within 90% confidence, whereas the free coefficient seems to not pass the test

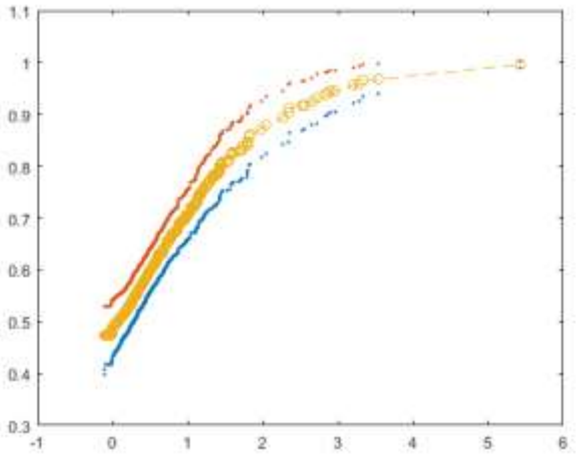

Figure 1: Probit regression for Increasing Expenditures after discounts

Therefore the utility of the attractiveness is obtained by probit regression as follows

$$Y^* = \{0.0698\} - 0.1367 * \text{GenderBuyer} + 0.0962 * \text{Age Group} + 0.0002 * \text{PhoneContact} + \varepsilon \quad (8)$$

From relation (3) we observe that the gender of buyers (F=1,M=2) is mostly decisive in the increasing number of visits in the market after sales, and usually male buyers are not more frequent in the market after discounts have been applied. The phone contact has a slight effect on it. The age group has comparable role to the gender of consumer. In Figure 4 is seen that the probability for more visits in the market is high for almost all the values of utility function (8) and only few values are less than 0.5. In this sense, for nearly all consumers' specifics, the marketing strategy (prices discounts) has been found attractive for peoples that respond by increasing the number visits in the market. Thus is the intermediate change on the consumer behavior. In the second stage, the final behavior is considered.

Now the response variable is the change in expenditure measured by the natural function "*is greater than*" e.g., in absence of the marketing stimulus. It is possible that the buyer, under budget constraint, would respond to the discount spending the same quantity of money and therefore just buying some more goods, so this variable is meaningful and not trivially known. Here the independent variables include even average expenditures before sales and registered cards consumer. The first is expected to give information about which consumer category has increased the expenses, and the second could inform the role of being a formal consumer. Performing probit regression we obtain the utility function

$$y^* = -0.916 + 0.093 * \text{C.Gender} - (0.0133 * \text{C.AgeGroup}) + 0.0199 * \text{AverageVisit} + 0.854\,\text{AverageExpenses} - 0.136 * \text{CardHolder} + (0.0013\text{Call}) + 0.0396\text{PeriodOpen} \quad (9)$$

In (7), the statistical significance is acceptable for all variables except Tel,Call and Age.Group (the p.Value is high, ~0.3) so we put it in parentheses. Notice that their coefficients are small and so this does not affect the estimation of the utility so we kept them in the equation (7). Now we make use of the binary outcome expression using continues probability. The switching value for utility is

$$P(\text{ExpencesAfter} > \text{ExpencesBeforeX}) = P(u \geq 0.08) \quad (9)$$

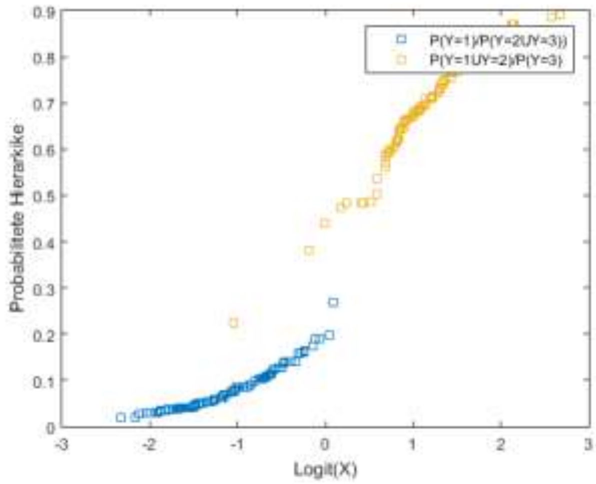

Figure 5: Full model probit regression

By using equation (8) we can realized conditions that (9) is fulfilled and therefore it is possible to forecast what happen with purchasing behavior for an individual with values

$$Consumer\{i\} \equiv n_1 n_2 \ldots n_7 \quad (10)$$

This is done by just putting values (10) in equation (9). We observe that rational or cognitive issues weight more in the utility value as seen from the coefficients for female byer that usually behave as major house holdings buyers, common average expenses that indicate the level of budget in the purpose, the time of effectiveness for sales. The expected psychic parameters as telephone call have less effect in shifting the utility. By randomly selecting consumer (predictor) values according to the population considered, we see that the value of (9) is usually reached, therefore the conclusion herein seem to be a global tendency in the area studied. Remember that those findings are general characteristics because the distribution in this state has been acknowledged as being stationary.

6. Conclusions

The harmonization of statistical techniques is an effective instrument in the analysis of econometric and psychometric systems. The study of changes as an argument of psychosocial elements with economic and financial ones requires high mathematical rigor for the influence of the utilization of different sectors of statistical analysis is a finding of value. From the modeling side, in the reviewed systems and others of these categories, the harmonization of statistics, numerical and analytical scaled factors are defined in the creation of a reliable calculation scheme. In the case of making purchases conditional on the customer's behavior being rational, he weighs the alternatives in terms of his use. In particular, to conclude that we consider the consumer towards discounts was characterized by the increase of the expenses themselves, not only of the volumes of the items bought. From a marketing point of view, this work suggests that the concessions have studied and programmed the program of carrying out their responsibilities not only to make the depreciation of a stock, but also for themselves the profit. We arrived at a more described finding of the system, in a more natural form of its other places.

References

[1] Claudio Castellano, Santo Fortunato,Vittorio Loreto.: Statistical physics of social dynamics. Rev. Mod. Phys. 81, 591-646 . April-June 2009

[2] Jisana T. K. Consumer behaviour models: an overview. Volume 1, Issue 5 (May, 2014)

[3] Jeff Bray. Consumer Behaviour Theory: Approaches and Models . http://eprints.bournemouth.ac.uk

[4] N. J. McCullen, M. V. Ivanchenko, V. D. Shalfeev W. F. Gale. A Dynamical Model of Decision-Making Behavior in a Network of Consumers with Applications to Energy Choices. http://www1.maths.leeds.ac.uk/

[5] Kadri G Yilmaz , Sedat Belbag. Prediction of Consumer Behavior Regarding Purchasing Remanufactured Products: A Logistics Regression Model. International Journal of Business and Social Research Volume 06, Issue 02, 2016

[6] Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. Statistics in Medicine, 22, 1433–1446.

[7] E.Kushta, D.Prenga, F.Memaj. International Journal of scientific research and management (IJSRM). Vol 6 N0 03(2018).

[8] Constantino Tsallis Computational applications of non-extensive statistical mechanics. Journal of Computational and Applied Mathematics 227 (2009) pp 51-58.

[9] G.P. Pavlos, M.N. Xenakis,L.P. Karakatsanis, A.C. Iliopoulos, A.E.G. Pavlos D.V. Sarafopoulos. Universality of Tsallis Non-Extensive Statistics and Fractal Dynamics for Complex Systems. Chaotic Modeling and Simulation (CMSIM) 2: 395-447, 2012.

[10] Sabir Umarov, Constantino Tsallis, Murray Gell-Mann, Stanly Steinberg. Generalization of symmetric -stable Lévy distributions for q>1. Journal of mathematical physics 51, 033502 2010

[11] Steiger, J.H. (1990), "Structural model evaluation and modification," Multivariate Behavioral Research, 25, 214-12.

[12] Sibiu Alma Mater University Journals. Series A. Economic Sciences – Volume 2, no. 4, December / 2009

[13] Horn, J. L. A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179-185.

[14] F. Mema. Marketing research in helping decision making" . "Ekonomia dhe Biznesi" Nr. 2 (22) 2006

[15] F. Mema. Statistical methods in acknowledgment of consumer behavior" International Review of Science, Innovation and New technology, Vol. 1, Nr 12, February, 2015

[16] Scott, David W. Multivariate Density Estimation and Visualization Papers Humboldt-Universitet Berlin, Center for Applied Statistics and Economics (CASE), no. 2004,16.

[17] Elmira Kushta, Teze doktorature

[18] Elmira Kushta, Dode Prenga, Fatmir Memaj. Analysis of consumer behavior in a small size market unit: case study for Vlora District, Albania. IJSRM,2018

[19] Mugo Fridah W.,Sampling in research.

[20] Jorge Faber and Lilian Martins Fonseca. How sample size influences research outcomes. Dental Press J Orthod. 2014 Jul-Aug; 19(4): 27–29.

[21] Sabir Umarov, Constantino Tsallis, Murray Gell-Mann, Stanly Steinberg. Generalization of symmetric -stable Lévy distributions for q>1. *Journal of mathematical physics* 51, 033502 2010

[22] M.A. Robinson. Quantitative research principles and methods for human-focused research in engineering design. '*Research methods' publications*. May 2016. DOI: 10.1007/978-3-319-33781-4_3

[23] Dolan, C. V. (1994). Faktor analysis of variables with 2,3,5, and 7 response categories: A comparizon of categorical variable estimators using simulated data. British Journal of Mathematical and statistical Psychology, 47, 309-326.

[24] Babakus, E., Ferguson, C., & Joreskog, K. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional asumptions. Journal of Marketing Research, 24, 222-228.

[25] DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. Structural Equation Modeling: A Multidisiplinary Journal, 9, 327-346.

[26] Barret, P. (2007) "Structural Equation Modelling: Adjudging Model Fit," Personality and Individual Differences, 42 (5), 815-24.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.