THE USE OF REGRESSION ANALYSIS IN MARKETING RESEARCH

DUMITRESCU Luigi
Lucian Blaga University of Sibiu, Romania

STANCIU Oana
Lucian Blaga University of Sibiu, Romania

TICHINDELEAN Mihai
Lucian Blaga University of Sibiu, Romania

VINEREAN Simona
Lucian Blaga University of Sibiu, Romania

Abstract:

The purpose of the paper is to illustrate the applicability of the linear multiple regression model within a marketing research based on primary, quantitative data. The theoretical background of the developed regression model is the value-chain concept of relationship marketing. In this sense, the authors presume that the outcome variable of the model, the monetary value of one purchase, depends on the clients' expectations regarding seven dimensions of the company's offer. The paper is structured in two parts. In the first part, a brief literature review enumerates the main multivariate data analysis methods used in marketing research and describes the general linear multiple regression model and its assumptions. The second part explains a set of procedures specific for the regression analysis.

Keywords: *Linear Multiple Regression, Marketing Research, Multivariate Data Analysis Methods, Relationship Marketing*

1. Introduction

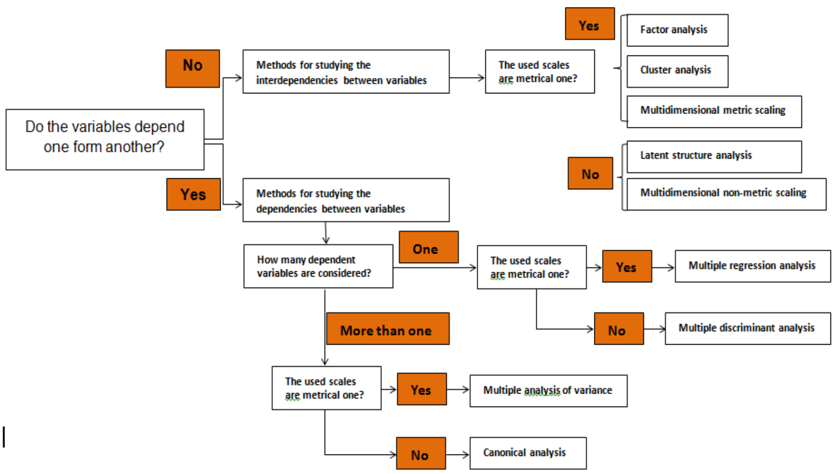
The core of decision making is information. Information is mainly the result of data analysis or data interpretation, thus it represents a result of applying different methods or techniques for creating value out of the gathered raw data. The instruments used to filter the raw data into a valuable piece of information differ according to the nature of the underlying data: qualitative or quantitative data. Qualitative data is unstructured data, collected through methods and techniques like: focus group interviews, depth interviews and projective techniques. Quantitative data is measurable data obtained through methods like: survey, observation, experiment and simulation. Measurement and scaling are the two main concepts used in defining and analyzing quantitative data. Thus, measurement represents the assignment of numbers or other symbols to characteristics of objects according to certain prespecified rules and scaling a generation of continuum upon which measured objects are located (Malhotra, 2010).

There are four main criteria upon which the methods used for data analysis differ within a marketing research: the used scale type (metric or non-metric scales), the number of researched samples (one, two or more samples), the relation between the researched samples (dependent or independent) and the number of used variables (one, two or more variables) (Catoiu, 2010). According to this classification, the multivariate data analysis techniques (techniques which use more than two variables) can differ if:

a. the considered variables are dependent one from another (causal relation) or if there is only a associative relation between them
b. there are one, two or more dependent variables within a causal relation
c. a metric scale or a non-metric scale is used for the considered variables.

The following figure contains the most known and used multivariate data analysis techniques:

Figure 1 – Multivariate Data Analysis Techniques



Source: Catoiu (2009, pg.552)

As seen in the figure above, multiple regression analysis is a technique that studies the dependency relation between one dependent variable and two or more independent variables, all measured on metric scales.

2. Literature Review

The combining of variables into a dependency relation is done by the economic theory in general and by the marketing theory in particular (e.g. one company's sales depend on the spent advertising money, different forms of price-bundles, the number of distribution channel and the size and quality of the general offer). Through empirical investigation, the theory is confirmed or not by the actual data or underlying behavior. One such instrument or method that confronts the theory with the actual behavior is the classical multiple linear regression model. This powerful instrument is used in various areas of economic research, such as international economics (Burnete, 2012) and macroeconomics (Opreana, 2010) and business research (Pop, 2010). The classical linear multiple regression model studies the linear relationship between one dependent variable and two or more independent variables.

The general form of the model is: $$Y = f(X_1, X_2, X_3, \dots, X_k) + \varepsilon$$
$$= X_1 \beta_1 + X_2 \beta_2 + \dots + X_k \beta_k + \varepsilon ,$$

where Y is the dependent or explained variable and X_j, with j=1,k are the independent or explanatory variables. The symbol ε represents the disturbance of the model or the residual variable. The set of independent variables explain to a certain extend the dependent variable; the rest of it is contained within the disturbance. The disturbance arises for several reasons, primary because it is not possible to capture every influence on a specific economic variable in a model, no matter how elaborate the model is and due to errors in measurement (Green, 2003).

Because of several constrains (time, cost, human resource, etc.), most of the marketing researches use statistical representative samples from which the data is collected. In this context, the linear multiple regression model has the following form:

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ik}\beta_k + \varepsilon ,$$

with i=1,n, where n is the used sample size.

The purpose of the model is to estimate the underlying unknown parameters of the model used further to check the validity of the stated theory and infer the sample's results on population level.

Because the regression model is based on two parts (a deterministic part: the independent variables and a random part: disturbance), the used data should respect some assumptions for proper model estimation and inference procedures (Green, 2003). The following section contains the six standard assumptions regarding the multiple regression model that will be explained through SPSS procedures in the research part of the paper.

According to Green (2003) the assumptions of the linear multiple regression model are:

1. Linearity: the model specifies a linear relationship between the outcome variable and the independent variables. Therefore, the mean value of the outcome variable for each increase of the predictors lies on a straight line.
2. Full rank (multicolliniarity): there is no exact linear relationship between any of the independent variables in the model.
3. Exogeneity of the independent variables: any estimated residual value is not a function of the independent variables. This means that the predictors do not contain useful information about the residual variable.
4. Homoscedasticity and nonautocorrelation: the residuals have a constant variance σ^2 within every level of the predictors and are not correlated one with each other.
5. Exogenously generated data: the process generating the data operates outside the model assumptions, thus independently of the process that generates the residuals.
6. Normal distribution: The residuals are normally distributed.

3. Research methodology

The purpose of the paper is to illustrate the application of the linear multiple regression model within a marketing research that uses primary, quantitative data as input. In this sense, five research objectives (O1 – O5) were developed accordingly to the assumptions of the linear multiple regression model.

Objective 1: To check if the developed model is a linear model.
Objective 2: To check the existence of multicollinearity between the predictors.
Objective 3: To check the independence of the residuals.
Objective 4: To check the homoscedasticity and non-autocorrelation of the residuals.
Objective 5: To check the normality of the residuals.

The developed regression model is based on the value chain concept within the relationship marketing theory (Bruhn, 2010). The company's outputs are considered the results of a complex interaction between consumer specific factors (observed or deduced by the company) which, themselves are generated by the company's actions. One of the deduced, customer specific dimensions is customer satisfaction. Customer satisfaction is defined by Oliver (1999) as a complex process of the information analysis. Bloemer (1998) describes the customers' satisfaction regarding a certain brand as an experience felt by them after consuming that brand, as a subjective evaluation of the way in which the brand's performance equals or not, performs or not to their initial expectations. Generally, satisfaction is the pleasure or disappointment felt by someone in the moment when he compares his initial expectations regarding the brand with its actual performance obtained through consumption (Kotler, 2009). The result of the evaluation process can be favorable for the client, if the brand's performance exceeds his initial expectations regarding that certain brand (the client is satisfied), or not, if the brand performance is under the initial expectation level (the client is unsatisfied or even annoyed, and feels abused). A latent or fragile satisfaction is felt if the brand performance equals the initial expectations. From the mentioned satisfaction theory, the authors assume that the dimensions which form the satisfaction regarding a company's offer influence the monetary value spent by the customer within his last commercial contact with the company. By considering satisfaction as a subjective evaluation process (simply said as a mental difference), a further presumption could consider that the initial expectations of a customer's regarding the company's offer (perceived through its dimensions) may influence in some measure the spent monetary value.

In our particular case, we have chosen a fuel distributing company which has one sell point in the town of Sibiu, Romania. Moreover, we have considered the following seven underlying dimensions of clients' satisfaction: time spent within the fuel company, fuel quality, personnel politeness, shop diversity, personnel response to the clients' problems, auxiliary services that can be used by the clients and personnel readiness to help the clients. The clients' expectations regarding the mentioned dimensions were measured through nine-point interval scales with the scale poles representing the minimum/maximum expectation for the considered dimension (table 1).

Table 1 – Example of scale used for measuring the expectation regarding the fuel quality

<table>
<tr><td rowspan="3">E1</td><td rowspan="3">Fuel quality</td><td colspan="9">Expected level of performance</td></tr>
<tr><td colspan="8">Low</td><td>High</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td></tr>
</table>

These seven variables are considered independent variables or predictors which influence one client's spent monetary value within his last contact with the fuel-company. The monetary value was measured in Lei (Romanian currency) using a ratio scale. As information source, end-customers of the fuel station were selected; they represent an external information source (clients who are part of the external environment of the fuel company), a primary information source (the obtained data is analyzed for the first time by the authors), and also they represent a free information source. A survey was used as research method and a questionnaire as research instrument. The sample size was represented by 10 respondents, for which the selection variables were two demographic variables (age [at least18 years old] and possession of a car) and a behavioral variable (people using the car for at least four times per week). The data collection period was from March 1 until April 8, 2011. The small sample size of 10 will surely affect the regression model consistency and assumptions, but, nevertheless, not accepting the model or rejecting some model assumptions don't dilute the value of the regression analysis.

The data was analyzed using the statistical software package SPSS V.19. Regression analysis requires a set of procedure which are automatically computed and displayed (through outputs) by the mentioned software.

A first output (table 2) of the analysis includes some descriptive statistics (mean and standard deviation) of the used variables (explained variable and explanatory variables).

Table 2 – Descriptive Statistics of the used variables

Variable name	Mean	Standard deviation	Number of cases
Monetary value (MV)	152	84,564	10
Time spent within the fuel-station (TS)	5,7	1,494	10
Fuel quality (FQ)	7,5	0,707	10
Personnel politeness (PP)	6,5	1,509	10
Shop diversity (SD)	4,9	1,912	10
Personnel response to the clients' problems (PR)	6,2	1,398	10
Auxiliary services that can be used by the clients (AS)	5,3	1,16	10
Personnel readiness to help the clients (PRead)	6,4	1,35	10

Source: own computation

The results reveal that a Rompetrol customer has spent a mean value of 152 Lei within his last commercial contact with the company. The variation of the monetary value variable (MV) is pretty high within the customer base, a value of $R^2 = 0,556$ indicates that the values are dispersed from the mean. Further information are contained by the high mean value (7,5) and low standard deviation (0,707) of the variable *fuel quality* (Q), which can be interpreted as a high level of expectations for fuel quality. A high level may be the result of good past experience of the customer with the company and, more important, of the intrinsic value of this central attribute – fuel. The expectations regarding the shop diversity have the lowest tolerance level (4,9), showing that this offer attribute (shop diversity) displays little importance for the customer.

A second output (table 3) contains the values of the Pearson correlation coefficient and their statistical significance for each pair of the model variables.

Table 3 – Correlation Matrix between the used Model Variables

Pearson Correlation	MV	TS	FQ	PP	SD	PR	AS	PRead
MV	1	0,058	0,186	-0,235	0,311	0,306	0,073	0,021
TS	0,058	1	0,053	0,517	-0,012	0,510	0,250	0,672
FQ	0,186	0,053	1	0,469	0,123	-0,112	-0,203	0,233
PP	-0,235	0,517	0,469	1	-0,096	0,158	0,413	0,545
SD	0,311	-0,012	0,123	-0,096	1	0,175	-0,386	-0,026
PR	0,306	0,510	-0,112	0,158	0,175	1	0,233	0,718
AS	0,073	0,250	-0,203	0,413	-0,386	0,233	1	0,483
PRead	0,021	0,672	0,233	0,545	-0,026	0,718	0,483	1
Sig. (1-tailed)	MV	TS	FQ	PP	SD	PR	AS	PRead
MV	.	0,437	0,304	0,257	0,191	0,195	0,421	0,477
TS	0,437	.	0,443	0,063	0,487	0,066	0,243	0,017
FQ	0,304	0,443	.	0,086	0,367	0,379	0,287	0,259
PP	0,257	0,063	0,086	.	0,396	0,331	0,118	0,051
SD	0,191	0,487	0,367	0,396	.	0,315	0,135	0,472
PR	0,195	0,066	0,379	0,331	0,315	.	0,259	0,010
AS	0,421	0,243	0,287	0,118	0,135	0,259	.	0,079
PRead	0,477	0,017	0,259	0,051	0,472	0,010	0,079	.

Source: own computation

The values within the correlation matrix are Pearson coefficients computed for each pair of the model variables. These values have no informational power if there is no statistical significance of the relationship they represent. It seems that there is *more* correlation between the explanatory variables then between the outcome and the predictors. Relative intense and statistical significant correlations (for a significance level of α=0,1) are found between PP-TS (0,517; p-value=0,063); PR-TS (0,51; p-value=0,066); PRead-TS (0,672; p-value=0,017); PP-FQ (0,468, p-value=0,086); PRead-PP (0,545, p-value=0,051); PRead-PR (0,718, p-value = 0,01) and PRead-AS (0,483, p-value=0,079). Although the theory suggests (Field, 2006) that substantial correlation between the predictors (multicolliniarity) might appear if the Pearson coefficient value is >0,9, the obtained results represent a first sign of **multicolliniarity** within the specified model,

The overall model summary is presented in the following table (table 4):

Table 4 – Overall Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
0,973	0,947	0,761	41,334	1,914

Source: own computation

The value of the multiple correlation coefficient R (0,973) certifies a strong linear correlation between the outcome variable (MV) and the predictors. Highly important is the coefficient of determination value (R Square = 0,947) which indicates the proportion of the outcome variable variance which is *determined* by the variation of the predictors. Thus, 94,7% of the variability in MV (monetary value) is influenced be the variability of the predictors, a great result which should awake a lot of question marks in every researcher's mind.

The adjusted R-Square value (0,761) gives us a clue regarding the possibility of generalizing the computed model from the used sample to the entire target population. This value should be as close as possible to the initial R-Square value (0,947). The difference between R-Square and adjusted R-Square (0,947 – 0,761= 0,186) is interpreted as the additional variance (18,6%) induced within the variance of the outcome variable by the sample model relative to a population model. This high percentage states that the obtained model is mostly valid for the used sample, its generalization will bias the outcome. The last idea is an assumption which can be confirmed or not by the results of a model validation process. In most cases, a model validation process presumes a training set and a test set, a presumption that makes not the purpose of the present paper.

The Durbin-Watson statistic indicates whether the residuals are independent one form another (**nonautocorrelation of the residuals**). A residual is the difference between the observed value and the predicted value, thus a correlation of every two residuals would mean a correlation of the underlying observations. An independence of residuals is observed if the Durbin-Watson value is near to 2. It is our case (D-W = 1,914), thus the assumption of residual autocorrelation is rejected.

Another output represents the ANOVA-table (table 5) which explains the variance of the outcome variable through the influence of the computed model and residuals.

Table 5 – ANOVA Statistics

ANOVA	Sum of Squares	df	Mean Square	F	Sig.
Regression	60943,004	7	8706,143	5,096	0,174
Residual	3416,996	2	1708,498		
Total	64360,000	9			

Source: own computation

The Total Sum of Squares determines the quantity of variance obtained if we have used the mean value of MV for making predictions. The Sum of Squares of the Residuals indicates the quantity of variance which is not explained or generated by the computed model. Thus, the Regression Sum of Squares stands for the variance of the predicted outcome variable generated by the computed model. All three variances are relative small, the main cause is represented by the reduced number of cases (10) upon which the regression model was built.

A high F-value shows the prediction consistency of the computed model relative to the residual influence. Its value is calculated as a proportion between the Regression Mean Square and Residual Mean Square (5,096). This F-value is positively related to the Regression and Residual Sum of Square, therefore high absolute Sum of Square values (more observation) and significant difference between Regression and Residuals Sum of Squares imply a high F-value. The statistical significance of 0,174 shows the probability that the computed F-value is obtained by chance. The present variance analysis reveals a high prediction consistency of the computed model relative to the residuals, but the small F-value indicates the lack of variance in the used variables (observed and predicted outcome variables) due to the reduces number of observations.

The estimated model coefficients and their statistical significance are presented in table 6:

Table 6 – Estimated Model Coefficients

	Unstandardized Coefficients		Standardized Coefficient	t	Sig.	95% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	-1126,5	252,799		-4,456	0,047	-2214,22	-38,805
TS	39,398	14,5	0,696	2,717	0,113	-22,993	101,788
FQ	149,47	31,07	1,250	4,81	0,041	15,763	283,18
PP	-56,16	15,085	-1,002	-3,723	0,065	-121,064	8,744
SD	13,725	8,24	0,31	1,666	0,238	-21,731	49,18
PR	57,104	17,656	0,944	3,234	0,084	-18,866	133,073
AS	83,814	19,302	1,149	4,342	0,049	0,763	166,865
PRead	-88,68	24,687	-1,416	-3,592	0,07	-194,901	17,541

Source: own computation

The model coefficients were estimated using the least squares method. If we include the estimated values, the regression line will have the following equation:

$$\widehat{MV}_i = -1126{,}5 + 39{,}398TS_i + 149{,}47FQ_i - 56{,}16PP_i + 13{,}725SD_i + 13{,}725PR_i + 83{,}814AS_i - 88{,}68PRead_i$$

B-values (estimated coefficients) represent one of the most important aspects of regression analysis. Their sign indicate the sense of the relation between the outcome variable and the associated predictor and their value the change which occurs in the outcome variable if the other predictors are constant or excluded from the analysis. Inverse relationships can be detected between the monetary value and personnel politeness and between monetary value and personnel readiness to solve the customers' problems. A simple interpretation of these findings is that low levels of expectations regarding the personnel politeness and readiness imply high levels of the spent value within the last contact with the company. It should be clear that the company's seven perceived attributes are measured through the customers' expectations regarding them. In this sense, the two identified inverse relationships may contain a clue that the attributes (personnel politeness and readiness) present little importance in the customer's decision making process. The assumption that the obtained results are pure mathematical findings that have no correspondence in reality should not be excluded.

T-values are computed values of the t-student test. They contain two kinds of information regarding the estimated coefficients: (1) their measure or magnitude indicate the contribution the associated predictor has on the model and (2) their statistical significance (Sig.) represents the probability that the estimated coefficient doesn't differ significant from 0. As expected, the fuel quality has a substantial influence on the outcome variable (t-value = 4,81). The statistical significance of the estimated coefficients is pretty good (all values are under the statistical coefficient of α=0,09) with two exceptions: time spent (Sig.=0,113) and shop diversity (Sig.=0,238).

A confidence interval was computed for each estimated coefficient. It is used to predict (with a certain probability) the true value of the estimated coefficients on population level. In our case, the probability that the estimated coefficient on population level is included in the computed interval is of 0,95. The theory suggests that the narrower the confidence interval is, the more precise the estimated coefficient on sample level is (Field, 2006). The data in the last two columns of table 6 present large confident intervals, therefore it is not recommended to use the estimated coefficients on sample level for predications at population level. Most of the confidence intervals (5 of 7) cross the point 0, indicating that in some samples the predictor has a negative relationship to the outcome variable whereas in others in has a positive relationship (Field, 2006). This case suggests a lack of consistency in the computed model. The small number of samples observations reveals, once again, the impossibility of inferring the sample results on population level.

Table 7 – Zero-order, Part and Partial Correlations

	Correlations		
	Zero-order	Partial	Part
(Constant)			
TS	0,058	0,887	0,443
FQ	0,186	0,959	0,784
PP	-0,235	-0,935	-0,607

SD	0,311	0,762	0,271
PR	0,306	0,916	0,527
AS	0,073	0,951	0,707
PRead	0,021	-0,930	-0,585

Source: own computation

The table above (table 7) contains the three sets of correlation coefficients computed by SPSS. Zero-order correlations are simple correlation between the outcome variable and each predictor. Their interpretation shows the fact that if we would use simple linear regressions between the outcome variable and each predictor, there would be none or a low correlation between the mentioned pairs of variables.

Partial correlations identify the pure or unique correlation between two variables (e.g. the outcome variable MV and a predictor TS) by excluding the effect (or the variance) of one or more control variables (e.g. FQ, PP, SD, PR, AS, PRead) on both initial variables. High values are obtained for the correlation between each pair of the outcome variable and a predictor, by eliminating the effects of the other predictors. A more in depth analysis of these results certifies that these high partial correlations can be explained by a minimum of three other arguments: (1) Interval scales were used to measure the seven explanatory variables which imply (2) relative small variation within the variables and (3) a sign of **multicolliniarity** can be detected.

Part correlations (semi-partial correlations) detect the association between two variables (e.g. the outcome variable MV and a predictor TS) by eliminating the effect of one or more variables (control variables) on one of the two initial variables (e.g. eliminate the effect of PP on MV, but not on TS). All the values are smaller relative to the partial correlation values because the mass of variation is *more* within the part correlation then the partial correlation. The interpretation of these values is similar to the partial correlation, with the exception that the common variance of the control variables and one of the initial variables is not eliminated.

Under the mentioned consideration, it can be concluded that high expectations regarding fuel quality, auxiliary services and personnel response exert a high influence on the outcome variable – monetary value.

Table 8 – Collinearity Statistics

Model	Collinearity statistics	
	Tolerance	VIF
(Constant)		
TS	0,404	2,474
FQ	0,393	2,544
PP	0,366	2,73
SD	0,765	1,308
PR	0,311	3,211

AS	0,379	2,639
PRead	0,171	5,85

Source: own computation

For every estimated coefficient, the variation inflation factor measures how much of this variation is *inflated* (overestimated) due to the existence of **multicolliniarity** between the predictors.

From its formula: $VIF_k = \frac{1}{1 - R_k^2}$, it can be concluded that big values of this inflation factor are determined by the existence of multicolliniarity between the predictors. The value of R_k^2 represents the determination coefficient computed for the multiple regression between the k^{th} predictor and the other remaining predictors. Table 8 contains big values of the VIF for all the 7 predictors, indicating the clear existence of **multicolliniarity** between the predictors.

The second collinearity indicator is the invers of the VIF. Tolerance below 0.2 indicate serious problem of multicolliniarity (Field, 2006).

Table 9 – Collinearity diagnostics

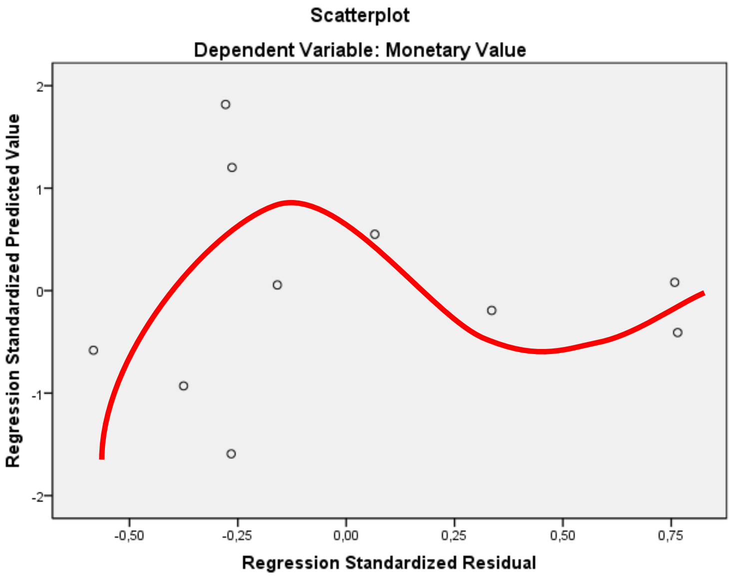
Dimension	Eigenvalue	Condition Index	Variance proportion							
			(Constant)	TS	FQ	PP	SD	PR	AS	PRead
1	7,735	1,000	,00	,00	,00	,00	,00	,00	,00	,00
2	0,135	7,558	,00	,01	,00	,01	,53	,00	,01	,00
3	0,048	12,654	,01	,13	,01	,02	,00	,07	,03	,01
4	0,036	14,726	,00	,14	,00	,16	,00	,08	,08	,00
5	0,021	19,150	,02	,00	,04	,06	,41	,02	,24	,00
6	0,016	22,010	,01	,49	,00	,22	,02	,10	,11	,06
7	0,007	32,974	,01	,00	,02	,31	,03	,48	,00	,50
8	0,001	78,133	,95	,22	,93	,22	,00	,25	,52	,43

Source: own computation

Further information regarding the possible existence of **multicollinearity** between the predictors is contained by the data within the above table. A lack of multicollinearity would mean that for each of the seven dimensions (we exclude the dimension associated with the constant) a high proportion of the variance of only one predictor should account for. It is not our case; multicollinearity of the predictors has once again been proven.

A final part of the regression analysis contains graphical representations based on which some other assumptions of the regression model can be checked (**linearity of the model**, **homoscedasticity of the residuals and normality of the residuals**).

Figure 2 – Scatterplot of the Regression Standardized Residuals and Regression Standardized Predicted Value

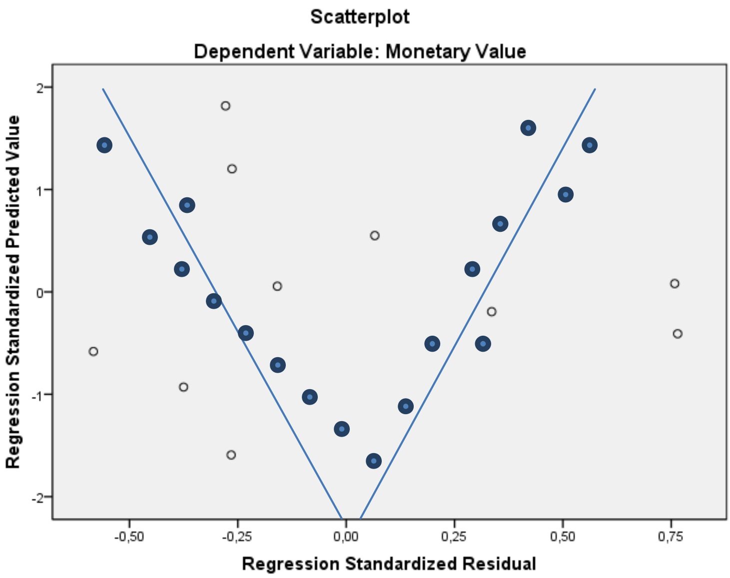


Source: own computation

Residuals are computed as the difference between the predicted value and the observed value for each observation of the database. The figure above overlaps the standardized values of the predicted values with the ones of the residuals. By analyzing it, the two assumption of **model linearity and homoscedasticity of the residuals** can be checked. For the first assumption to be confirmed (**model linearity**), the dots of the scatterplot should converge to a line. Their arrangement follows rather a non-linear path (denoted in the figure 2 by the red curve), signaling for the non-linearity of the model.

The second assumption of the **residuals' homoscedasticity** is confirmed if the dots of the scatterplot are randomly distributed in the bi-dimensional space. Homoscedasticity of the residuals should be understood as a constant variance of the residuals within each level of the predictors. A non-constant variance of the residuals would be graphically represented as a funnel which could display an increase of the residuals' variance related to an increase of the predicted values (Figure 3 – the displayed blue lines and dots are a typical example of heteroscedasticity).

Figure 3 – An example of heteroscedasticity of the residuals



Source: own computation

The distribution of the dots in figure 2 has no particular pattern, the dots are randomly distributed, and thus the assumption of constant variance of the residuals (**homoscedasticity**) is confirmed.

The next set of figures contains information regarding the assumption of the **residuals' normality**.

Figure 4 – Histogram of the Standardized Residuals

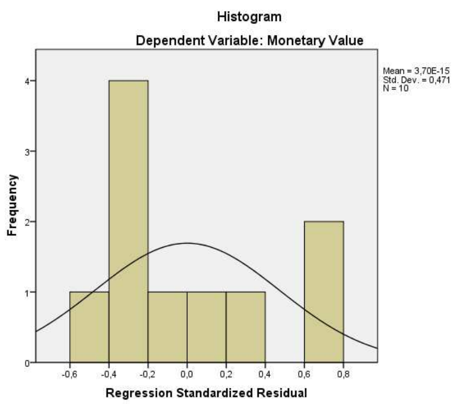


Figure 5 – P-P Plot of the Standardized Residuals

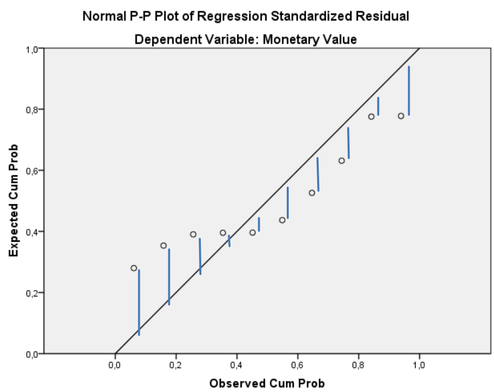


Figure 4 represents the histogram of the obtained residuals and the normality curve. It can be easily observed that the residuals do not follow a normal distribution; peaks that differ from the mean value can be observed for the intervals [-0,4 -0,2] and [0,6 0,8]. The same conclusion is drawn after analyzing figure 5. The vertical segments between the dots and the normality line indicate the absence of a normal distribution for the observed residuals. If the residuals were normally distributed; the dots would be identical with the normality line or would vary slightly from it. Thus, the normality assumption of the residuals is not confirmed.

4. Conclusions

The purpose of the present paper is to illustrate how to apply and interpret the linear multiple regression analysis within a marketing research based on primary, quantitative data. The model is based on the value-chain concept of relationship marketing theory. In this sense, the monetary value spent by a customer within his last purchase (dependent variable) is influenced by his expectations regarding the dimensions of the company's offer. The fuel-station offer was taken as study-object and the expectation for seven dimensions were measured through a nine-point interval scale.

After applying a set of regression specific procedures (using the statistical software SPSS v.19), two of the five regression assumptions were confirmed. Thus, the developed regression model is not a linear model, signs of multicolliniarity within the predictors are in every part of the analysis and the residuals do not follow a normal distribution. The computed residual values present constant variance at every level of the predictors (homoscedasticity) and do not correlate between each other (nonautocorrelation).

These results demonstrate the lack of consistency of the developed model. Several factors could have biased the desired results. All predictors are basically dimensions that form the company's offer. These dimensions are semantically and practically different; they can satisfy different customer needs. If a customer is asked to evaluate (on a nine-point interval scale) his expectations regarding the mentioned dimensions, a process of *mental smoothing* can appear because of everyone's desire of high expectation in general and regarding a company's offer in particular. The small variation within and between the predictors is the main cause for multicolliniarity. Another possible cause for the obtained results is the small sample size. Although the theoretical rule states that $n \geq k$ (number of observations should be greater than the number of predictors), the sample size of $n=10$ observations generates small variations within the variable, diminishing the prediction power of the computed model relative to the residuals. Thus, measurement and scaling techniques require essential improvement.

Further research will be orientated towards the applicability of other multivariate analysis methods and models (factor analysis, principle components, MANOVA, etc.) within the relationship marketing theory.

5. References

Bloemer, J., de Ruyter, K., 1998. On the relationship between store image, store satisfaction and store loyalty, European Journal of Marketing, Vol.32, No.5/6, 499-513

Bruhn M., 2010. Relationship Marketing: Das Management von Kundenbeziehungen, München, VahlenVerlag

Burnete S., 2012. New EU Member Countries are Phasing out Labor-intensive Activities – an Econometric Approach, Working Paper

Cătoiu I. (Coordinator), 2009. Tratat de Cercetări de Marketing. București. Editura Uranus

Field A., 2006. Discovering Statistics using SPSS. Second Edition. London. Sage Publications.

Greene W.H., 2002. Econometric Analysis. Fifth Edition. New Jersey. Pretince Hall.

Kotler Ph., Keller J., 2009. Managementul Marketingului, Ediția a V-a, București, Editura Teora

Malhotra N., 2010. Marketing Research. An Applied Approuch. Sixth Edition. Ney Jersey. Pearson Eduction.

Neagoe C., 2012

Oliver, R.L., Swan, J.E., 1989. Customer Perceptions of Interpersonal Equity and Satisfaction in Transactions: A Field Survey Approuch, Journal of Marketing, Vol.53, 21-35

Opreana, A., 2010. The Long-Run Determinants of Investment: A Dynamic Approach for the Future Economic Policies, Studies in Business and Economics, Vol.5, Issue 3, 227-237.

Pop N. Al., Neagoe C., Vlădoi A.-D., 2010. Comunicarea integrată în marketingul intern. Studiu de caz pentru întreprinderile mici și mijlocii. Proceedings of the 5th International Conference on Business Excelence, Vol. 2, 89 - 92.