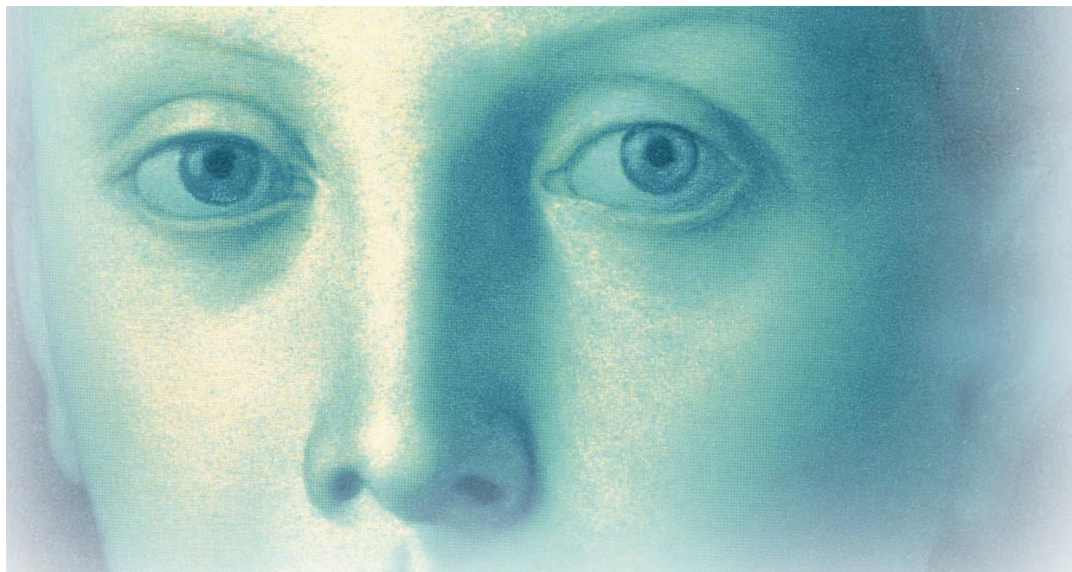




White Paper



The derived importance of attributes and the problem of collinearity

M.V.Bodunov
Senior analyst
Ipsos Russia



In market research the evaluation of the importance of the properties of the objects (products, concepts etc) is often used by relating the corresponding variables with several "external" criteria (for example general evaluation of the concept or the intention to buy the product). If any variable is closely related with a criterion, one can speak about the high importance of this variable, and vice versa, if the variable is loosely bound with the criterion, it witnesses of its low influence and importance. The importance, measured with the use of such method is called the derived (or analytically derived) importance [Tsisar' A., 2001]. It is assumed that the obtained values of the importance with the use of such indirect approaches are less depended on the personal control from the respondent side than the claimed evaluations of the importance of the examined characteristics.

As a rule the correlation and regression analysis is used to evaluate the relations between the examined variables and the criterion. At that the result is strongly depends on how closely the independent variables under the analysis of the importance are coupled with each other. In ideal situation that is rare in practice, the tested variables do not correlate with each other (i.e. are mutually orthogonal). In this case the corresponding correlation and regression coefficients are coincide with each other, and their sum of squares will be equal to the explained variance. The square of correlation (regression) coefficient taken as portion of the explained dispersion will provide the value of relative (variance) importance of the corresponding variable.¹ Only in this case the relative importance of different variables won't depend on each other (naturally in case of constant set of predictors as it determines the full explained variance).

On the contrary, in case when the moderate or strong relations can be seen between the examined variables, the situation dramatically changes. The corresponding correlation or regression coefficients won't coincide and they acquire quite different meaning. The coefficient of correlation reflects the specific contribution of each examined variable (and this contribution is composed of two main components: the unique contribution of the given variable that is not related with other variables included into the analysis and of the general component related with the correlating variables). In contrast to the correlation coefficients, the regression coefficients show which additional contribution in the explanation of the dependent variable is taken by each predictor in the combination with other predictors included into the analysis. In the presence of collinearity the coefficients of correlation can be used as a measure of contribution into the explanation of the criterion, if the researcher is interested in specific contribution of each separate predictor without the examination of which and in what portions the latent variables participate in the determination of the variability of these predictors. It is impossible to speak about the importance of the predictors on the basis of correlations themselves. The situation with the collinearity can be implied even more negatively on the regression coefficients. The collinearity may results in significant underestimation of the contribution of highly correlated variables. Sometimes the collinearity can be revealed in the appearance of the negative regression coefficients for the variables, which has the only positive relation with the criterion.

At the present time a number of methods and procedure are used to improve the evaluation of the importance by elimination or compensation of the collinearity effect.

One of the methods (especially efficient in the cases when the number of

¹ The coefficients (correlation and regression) are called the indicators of the theoretical importance. The notion of level importance is also used, which is equal to the increase of the mean value of the criterion at the increase of the corresponding predictor by one unit. This measure is very popular in economics [Johnson, J. W. et al, 2004].

the independent variables is great) is the use of the factor analysis applied to the examined variables and the transformation of the original set of variables into the new ones, mutually non-correlated (orthogonal) indices.

The following methods are based on the regression analysis that is applied to the original set of the examined variables without their preliminary factorization:

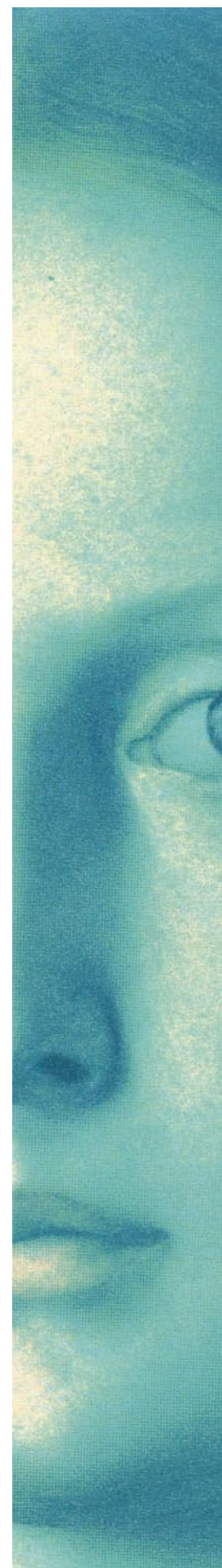
1. Ridge regression.
2. Methods that examine all possible combinations of the independent variables.
 - a. Relative importance by Kruskal;
 - b. Shapley regression indices;
3. Relative importance under Johnson.

The relative importance is determined as a proportional contribution of each independent variable into R^2 , which takes into account its direct effect (i.e. its correlation with the criterion), as well as the effect in combination with other variables in the regression equation (Johnson, 2004, c. 240).

If we give the stricter classification of the approaches than we can name three groups:

1. The reduction of dimension by means of the methods of grouping correlated independent variables with the creation of latent variables (for example, regression models on principal components). Unfortunately these approaches don't let to interpret (isolate, evaluate) the effect of the separate original variables other things being equal. From the other hand this approach provides the evaluation of the inner dimensionality of the original data and to move the analysis to the level of more consistent generalized variables (in case of their acceptable content interpretation).
2. The correction of the coefficients with the use of scalar parameter of proportional shrinkage (dimensional scaling), for example, ridge regression. The disadvantage of this approach is the fact, that the addition of the scalar quantity to the elements of the main diagonal of the covariance matrix diagonal does not improve the interpretation. In addition there are no clear criteria for the choice of the value of the scaling parameter.
3. The transformation of the independent variables with the purpose to remove the correlation. For example orthogonalization of variables-predictors that is carried out with the use of the sequential regression (Kruskal, 1987). Such approach keeps the possibility to interpret the independent variables. From the other hand the coefficients turn to be strictly depended on the order of their inclusion into the model. To eliminate such dependence one can use the technique of averaging of contributions of each variable in all possible combinations with other independent variables. However if the number of the predictors is large one can hardly perform the analysis with the available software. The significant advantage of the relative weights (as indices of the importance) is that they are more stable from one research to another, than the standard regression coefficients. In addition, in contrast to the regression coefficients, the sum of all relative weights is equal to the squared multiple correlation (R^2).

Thus the regression coefficients in the presence of correlation between the predictors turn out to be bad indices of influence (contribution, weight,





importance). From the other hand there are no any alternative methods of the evaluation of their importance.

The acceptable results can be obtained with the use of ridge regression. The optimal technique is the method of relative weight evaluation (averaging over orderings). There is a heuristic approximation of the method of average on all the combinations, proposed by J. Johnson (Johnson, 2004).

The relative weight (importance) evaluation method includes the following steps:

1. The criterion variable is selected to fully reflect the general likeability, performance, or effectiveness of the examined object.
2. The potential predictors or drivers of performance or likeability are determined.
3. The relative importance of the selected sets of the predictors is evaluated. The relative importance can be expressed in percentage of the total variance of the criterion explained by each predictor.

The evaluation of the approximate values of the relative weight (importance) is carried out with the following procedure:

1. The multiple regression analysis is carried out with the selected criterion as a dependent variable, and the selected predictors as independent variables. At this stage the value of R^2 is registered.
2. The method of the principal components is applied to correlation matrix between the predictors. The number of the selected components is equal to the number of predictors (i.e. the maximum number of the component that explains the 100% of variance is extracted). The rotation of the components is carried out with the help of the Varimax or Equamax procedures. After the rotation each component is turned out to be loaded with the large loading by only one examined variable. The evaluation of the factor scores of the given component is carried out.
3. At the final stage the calculation of «orthogonalized» regression coefficients is carried out. First the coefficients between the factor scores and original predictors as well as between the factor scores and criteria are calculated. Than these two matrices (the square matrix and the vector) are multiplied together. The resulting vector contains the standardized relative weights (importance coefficients) of the predictors. As a matter of convenience each coefficient can be expressed as the percentage of R^2 (as their total sum is equal to R^2).

To illustrate the relations between the different indices of importance there are two examples below of their usage in typical situations of market researches. The first example relates to the testing of food product. The respondent's intention to purchase the product (5-point scale) is used as a criterion (dependent variable). The evaluation of the importance was carried out for the following integral characteristics of the respondent's perception of the grocery characteristics (9-point scale): taste, texture, and quality. The properties were evaluated after the respondent tasted the product. The intention to purchase the product was evaluated at the end. The Table 1 contains correlations (Pearson) of the indices of the properties of the product with the criterion (intention to purchase the product), as well as coefficients of linear regression, coefficients of ridge regression and the relative weights (importance under Johnson) for the given indices. By all the indices of the importance the maximum weight was found for the taste of

the product, and the minimum was revealed for the texture of the product. Such a coincidence of the rank position of the different indices of the importance is not met always (if the number of indices is large, one can often see the mismatch of the ranks, especially in the range of medium importance). The substantive differences between the corresponding coefficients of correlation and the coefficients of ordinary linear regression, as well as unrealistic negative regression coefficient for the texture signify the presence of the significant multicollinearity caused by the close relations among the independent variables (indices of the product properties). In this case (rather typical) it is not reasonable to use the values of correlation and regression coefficients as the indices of the importance (the effects of different product properties in them turn out to be non-separated). The ridge regression improves the balance between the values of weight for different indices, however it leads to the significant underestimation of these weights. The relative weights as the indices of the importance under Johnson are the optimal from the point of the balance, as well as from the point of the absolute values.

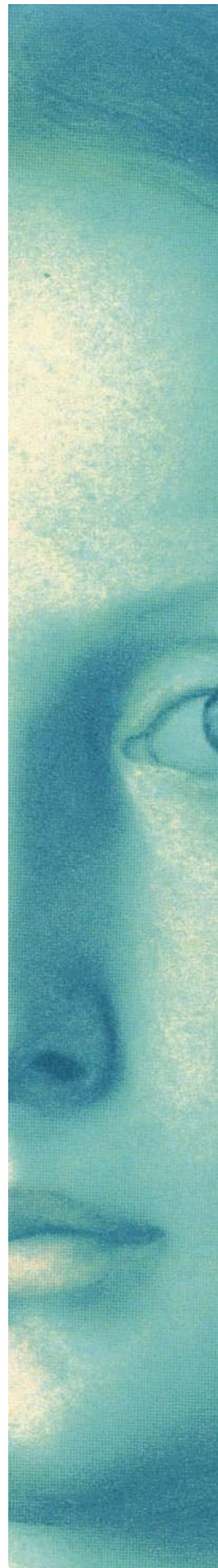
Table 1

Indices of product properties	Coefficients of correlation	Standardized coefficients of ordinary regression	Standardized coefficients of the ridge regression	Relative weights
Taste (Aroma)	0.800	0.606	0.287	0.351
Texture	0.544	-0.012	0.111	0.101
Quality	0.715	0.269	0.211	0.219

N = 850

R² = 0.671

The second example presents the results of testing of the package for the food product. The estimation of the respondent's intention to purchase the product (7-point scale) was used as a criterion (dependent variable). Different characteristics of package (5-point scale) for which the estimation of the derived consumer importance was carried out, were used as the independent variables. The total number of the predictors was 19, i.e. significantly more than in the first example. The factor analysis of the relations between them was carried out before the estimation of the importance of these characteristics (Principal components, Varimax rotation). It is advisable to carry out the factorization in all cases when the estimation of the derived importance is performed for a large number of predictors at the same time (more than 10). Thus the following is achieved: the real inner dimensionality of the examined set of the characteristics becomes clear; the additional important information for interpretation of the relationships between them is obtained. In the number of works that examine the methods of the elimination of collinearity, the factorization is proposed as one of such techniques of orthogonalization of the predictors. Though such method is fully justified, it is noted that its usage can result in the loss of information on the unique contribution of the predictors into the variation of criterion (i.e. the information on the importance of the predictors with the unique components taken into account). If the researchers are interested in the contribution of the general aspects of the examined indices, and they are not so interested in the unique contribution of the separate





predictors, than the usage of the general factors providing their good interpretation becomes the best solution of the problem of the multicollinearity (except for the cases with small samples). In this case the coefficients of correlation of the predictors with the criterion will be equal to the corresponding regression coefficients, and their sum of squares will give the explained variance (R^2).

In case of our 19 indices by the method of principal components the two general factors were extracted, that jointly explain 62.4% of the total variance. The factors after rotation (presented in the Table 2) are easily interpretable: the first factor is related to motivating function of the package (persuasion), while the second factor reflects mainly the perceptual peculiarities of the package, its memorability. Some of the predictors turned out to be polymetric, i.e. correlated with the both factors (for example «Suitable for an expensive product», «Looks modern»,) that can show the necessity to make the component structure more exact.² However the given structure seems satisfactory for our illustrative purposes.

Table 2

Packing properties	F1 (34.3%)	F2 (28.1%)
Is suitable for an expensive product	0.498	0.447
Makes me want to try it	0.620	0.480
Memorable pack		0.707
Looks cheap		-0.619
Looks modern	0.475	0.527
Good to serve to guests	0.661	
Stands out on a shelf		0.757
Bright, eye-catching design		0.747
Presents a trustworthy brand	0.722	
Stylish	0.602	0.539
Dull and boring		-0.704
Communicates great taste	0.660	0.493
For people like me	0.770	
Good value for money	0.825	
Is a pack for natural product	0.788	
Is a pack for a product that is worth paying a bit more	0.653	
Is a pack for a product that tastes like freshly prepared	0.740	
Has a unique and original design		0.668
Fits the image of a premium product	0.625	0.593

The correlation coefficients of those factors with a criterion (purchase intention) are 0.566 и 0.418 respectively (ordinary regression coefficients have the same value because of factor orthogonality). The squares of the indicators give the following values of the derived importance as 0.320 и 0.175 (with $R^2=0.495$ – 64.7% and 35.3%). Thus the first factor (“Motivating role of packing”) being a bit more powerful was twice as more important than the factor of the visual and perceptive packing features.

Let’s consider the indicators of the resulted importance obtained for all 19 indicators (Table 3). I would like to note that taking into account the unique contribution of the analyzed indicators increased the explained variance of the criteria by 10% from 0.495 in factor model to 0.545 in the detailed model. Like in the first example the strong multicollinearity (mostly reflected

² The purpose of the given article is not to examine in detail the problems related to the usage of the factor analysis in the market research. This theme will be discussed in the special paper.

in the factor structure of predictors) led to substantive difference between the correlation coefficients and regression coefficients as well as to many sign inversions in the last ones (in both directions, see coefficients for «Looks cheap»). The usage of the ridge regression improves the picture in the presence of the strong decrease of the general level of the coefficients (for their transform into the indicators of importance they are to be squared except the 4th column of the Table 3 - Relative weights). The Johnson's values of relative weight have optimal properties.

Table 3

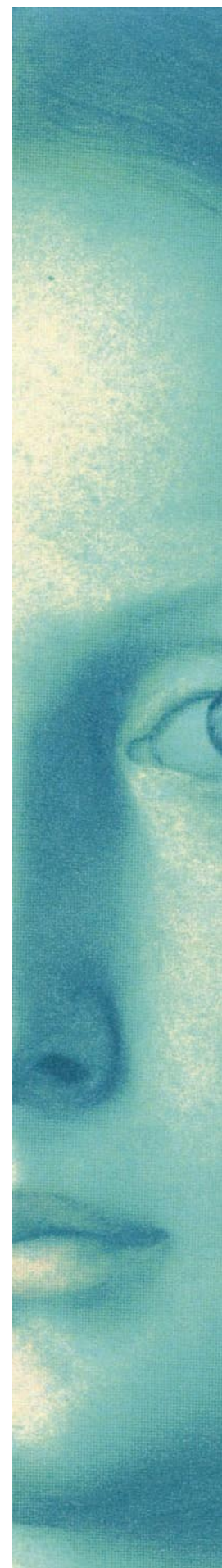
Product properties	Correlation indicators	Standard indicators of the usual regression	Standard indicators of the ridge regression	Relative weights
Makes me want to try it	0.631	0.233	0.103	0.067
Fits the image of a premium product	0.642	0.197	0.084	0.050
Memorable pack	0.568	0.165	0.080	0.045
Good value for money	0.618	0.196	0.080	0.045
For people like me	0.615	0.099	0.073	0.041
Is a pack for natural product	0.553	0.100	0.069	0.037
Stylish	0.554	0.048	0.046	0.030
Is a pack for a product that tastes like freshly prepared	0.558	0.022	0.049	0.029
Has a unique and original design	0.485	0.056	0.051	0.027
Is suitable for an expensive product	0.463	0.056	0.041	0.023
Is a pack for a product that is worth paying a bit more	0.490	-0.028	0.027	0.022
Communicates great taste	0.526	-0.056	0.016	0.020
Presents a trustworthy brand	0.467	-0.078	0.015	0.019
Stands out on a shelf	0.489	-0.084	0.014	0.018
Bright, eye-catching design	0.475	0.000	0.020	0.017
Good to serve to guests	0.463	-0.026	0.016	0.016
Dull and boring	-0.391	-0.025	-0.021	0.015
Looks modern	0.433	-0.006	0.024	0.014
Looks cheap	-0.359	0.055	0.001	0.010

N = 300

R² = 0.545

The data in Table are presented in the decreasing order of the relative weights (column 4).

The indicators with the highest importance (in that example – “Makes me





want to try it”, «Fits the image of a premium product», “Memorable pack”, “Good value for money”, “For people like me”) can be marked as key drivers.

In practical market research the relative importance of the properties, characteristics and attributes of the products are very often considered separately, out of any context. In theoretical research the evaluation of the attribute importance is always included into any system of notions or concepts, in frame of which the consumer behavior is analyzed and explained.

Let’s consider some aspects of the modern ideas about the correspondence between the derived importance of attributes of consumer objects, from one side, and the general characteristics of the consumer behavior as an integral system, from another. The understanding of that correspondence can assist more effective strategies for the positioning and promotion of the products/goods/services in the market conditions at the present time to be developed.

The importance of the product attributes (and its other characteristic) for the person can be associated with the personal characteristics of different levels: from the unique features of the individual experience to such general parameters as personal values. In general situation such relation is to be set up at the subconscious level due to the habitual actions. This relation is reflected (and formed) in the different preferences conditioning choices in consumer behavior continuum. If this relation is realized on the conscious level its influence can achieve the highest personal regulators of the individual behavior – general personal motives and values. Due to this relation the person can satisfy the need in the internal consistency, coherence of his/her behavior. The most important is that the actions consistent with the personal value preferences cannot require any external reinforcement because such actions themselves give to the person what he/she wants. Because of their abstractedness (ideality) the potential non-achievement of goals conditioned by the value preferences does not affect the person destructively. On the contrary such frustrating (abstract) goals can become more attractive for us. Their relative independence of situation provides the person with a plenty of the ways to coordinate the behavior acts (in particular the acts of the consumer choices) with the person’s value structure. All of that makes the values very important and perspective object (and instrument) in the market research.

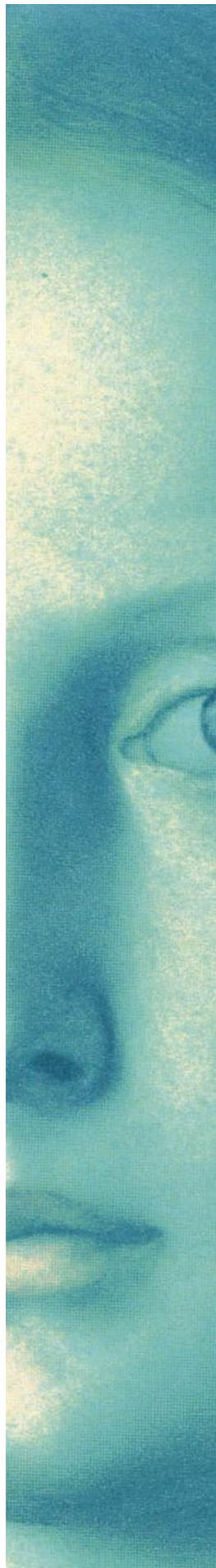
Nowadays several methods of diagnostic of the value structure of the person (consumer)³ have been developed. The Rokeach Value Survey, the Schwartz Value Survey and the Kahle’s List of Values – LOV are the most popular ones. The last one was specially developed for research in the consumer behavior.

In conclusion we consider the hypothetical interpretations of the results shown in the examples (Table 1 and 3) in terms of consumer value preferences. “Taste” attribute was of the first importance and “Quality” attribute was of the second importance in the first example. The first attribute is connected with such a value as hedonism, getting pleasure enjoyment from the life. The second attribute is to be related with the “Self-respect” as a value. Perhaps the balance of those two attributes indirectly reflects the dominating role of the “Hedonism” in comparison with “Self-respect” in the individual consumption of the FMCG food products. In the second example such packing feature as “Makes me want to try it”

³ That article does not have a task to present the problem of values in details, their measurement and usage in the market research. It will be the subject of one of the next articles.

achieved the highest importance, which is also related to the hedonistic value preferences. The second place according to the importance belongs to the feature “Fits the image of a premium product” relevant to the value “Being well-respected by others”. It can be supposed that those two universal values – the hedonism and respect by others– are the leading in the evaluation of the packing for the FMCG. The existence of two segments with a prevalence of one of these two value preferences can be supposed.

Thus the examination of the attribute importance in the view of the value preferences of the consumers (even on the hypothetical level) increases the possibility to interpret the results obtained and understand the mechanisms of attributes hierarchy according to their importance.





References

1. Beatty, S. E., Kahle, L. R., Homer, P.M. and Mirsa, S. Alternative Measurement Approaches to Consumer Values: The List of Values and the Rokeach Value Survey. *Psychology and Marketing*, Vol. 2 No. 3, 1985, 181-200.
2. Frost, Alan. A Consumer Choice Simulation Model for the Development of Product, their Packaging and Advertising. In Ian Greig (Ed.) *Brand Choice Modelling*, ESOMAR Monograph No 5, 1998, 7-26.
3. Johnson, Jeff W., LeBreton, James M. History and Use of Relative Importance Indices in Organizational Research. *Organizational Research Methods*, Vol. 7 No. 3, July 2004, 238-257.
4. Kruskal, W. Relative importance by averaging over orderings. *The American Statistician*, 41, 1987, 6–10.
5. LeBreton, James M., Binning, John F., Adorno, Anthony J., Melcher, Kathleen M. Importance of Personality and Job-Specific Affect for Predicting Job Attitudes and Withdrawal Behavior. *Organizational Research Methods*, Vol. 7 No. 3, July 2004, 300-325.
6. Overby, Jeffrey W., Woodruff, Robert B., Gardial, Sarah F. The influence of culture upon consumers' desired value perceptions: A research agenda. *Marketing Theory*, Vol. 5 No. 2, 139–163.
7. Tsisar' A. Review of the methods to measure the meaning of the good / service attributes (In Russian). *Маркетинг и маркетинговые исследования (Marketing and market research)*, №5, 2001
8. Weiner, Jay L., Tang, Jane. Multicollinearity in Customer Satisfaction Research. White Paper. Ipsos Loyalty, 2005.
9. Zeithaml, Valarie A. Consumer Perception of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. *J. of Marketing*, Vol. 52 No. 3, July 1988, 2-22.