UKD 330.014

M. Szłapka
Silesian School of Economics and Languages in Katowice, Poland

QUANTITATIVE METHODS IN THE PROCESS OF SEGMENTATION

У статті наведено основні припущення щодо здійснення сегментації клієнтів і кількісні методи, що використовуються при цьому. Визначено методи перевірки статистичних гіпотез для визначення статистичної значущості досліджуваних параметрів. Для того щоб уникнути дублювання інформації та спростити розподіл одержувачів на однорідні групи визначено, що може бути використаний аналіз співвідношення параметрів. При визначенні однорідних класів - сегментів покупців, застосовуються методи ієрархічного групування і розподілу. Незалежно від ієрархічних методів застосовуються також графічні методи, методи площі та щільності, оптимізація ітераційних методів та спільний аналіз.

In the article the fundamental assumptions of conducting the process of segmentation of customers and quantitative methods helpful in its realization are presented. The methods of the verification of statistical hypotheses enable determining the statistical relevance of the studied parameters. In order to avoid the redundancy of the information and to simplify the distribution of recipients into homogeneous groups, the analysis of the correlation of parameters can be used. In defining homogeneous classes - segments of buyers, the methods of hierarchical grouping and distribution are applied. Irrespective of hierarchical methods, the graphic, area and density methods, as well as iterative optimization and conjoint analysis are also applied.

Determining the scientific problem. In conditions of the intensive competitive policy it has been widely believed that gaining an advantage, which underlies every success, requires a lot of penetrating analyses and the ability of noticing the diversity of needs and preferences of buyers ([4]). The process of dividing the customers, in the range of the market, into groups, i.e. sections of similar or the same requirements that can be fulfilled by a specific marketing composition, is called segmentation ([5] p. 34). The most frequent problems associated with this issue are among others determining the manner of distinguishing the sections and the choice of the ones that give the greatest chances for the activity of the company.

The number of possible variables of segmentation is theoretically unlimited, however its high level makes the segmentation procedure difficult, morover it is recommended that data acquired about customers are essential for the enterprise and do not copy the same information ([5] p. 93, 143). The complexity of the society and the multitude of consumers' criteria means the necessity to select these parameters which will decide about the success of the enterprise. In these actions the methods of multicriteria analyses and examining the correlation of parameters can be helpful.

Analysis of the latest research and publications. The notion of *segmentation* was in classical terms ([3], [4], [5]) defined in literature as a division of the market into relatively homogeneous classes of buyers who are characterised by the same or similar requirements fulfilled by a specific marketing composition.

M. Walesiak ([8] p. 22) distinguishes two basic criteria of distribution of buyers of consumer goods:

- criteria characterizing the buyers:
  - geographical criteria: territorial range, size of the region, size of the city, population density, climate,
  - demographic: age, sex, family size, development phase of the family, religion, nationality, race,
  - socioeconomic: profit, education, profession, social class, lifestyle, personality,
- criteria characterizing consumers' behaviours in the area of:
  - reasons of the purchase: permanent and special,
  - expected benefits concerning the quality, services, savings,
  - the use: status of the user, speed of the use,
  - the attitude towards the product: enthusiastic, positive, indifferent, negative, hostile.

The same author distinguishes two basic types of segmentation according to its accompanying market:

- segmentation of buyers on the market of goods and consumer services (*consumer market segmentation*),
- segmentation of buyers on the market of goods and production services (*industrial market segmentation*).

Because of the specificity of the industrial market, criteria adopted on it and the course of proceedings are a little bit different. One of the recommended attempts in this area is a procedure of the two-stage segmentation into macro- and microsegments. Macrosegments are distinguished by the criteria characteristic of an enterprise-buyer: geographical, technical-economic and organizational parameters. Microsegmentation is conducted only when the results of macrosegmentation turn out to be unsatisfactory and its base are the parameters of the individuals purchasing from the enterprise-buyer, such as demographic parameters, personality, style and motives of making the decision about the purchase, attitude towards sellers, tendency to take a risk, etc.

From a wider perspective the main object of the research is not a single consumer but a whole community on a specific market e.g. the workers of one profession, residents of one city, etc. From this point of view the aim of segmentation research is an identification of classes of purchasers sharing certain characteristic parameters. Such a segmentation is conducted in two phases. In the first, on the basis of criteria characteristic of communities of buyers, macrosegments are distinguished. In the second phase, on the basis of characterizations of single buyers, microsegments are formed.

Examples of conducted segmentations include works of English-speaking authors from the 1980s - Saunders, Lessig, Anderson, Cox and others[1]. In 1990, S. Mynarski (Polish literature) conducted the segmentation of consumers on the furniture market.

Quantitative methods have been applied in the market segmentation for many years. Because of the need to consider many criteria of distinguishing the segments, most often in this process there are used a number of multidimensional analysis methods, discussed in detail in positions [2] and [7]. Apart from the most often enumerated methods of classification, those cited in positions [1] and [8] can be helpful as well.

The aims of the article. The purpose of the article is presenting quantitative methods useful in the process of segmentation of buyers and introducing their possible application with special attention paid to the analysis of the correlation of parameters as a tool compatible with assumptions and aims of the segmentation process, which enables determining the best criteria of distributing the recipients into homogeneous groups.

The basic scientific material of the article. *Segmentation process and possible implementations of quantitative methods.*

The most important problem of segmentation is selecting its appropriate criteria, bearing in mind that each of the conducted segmentations varies in time.

[1] Walesiak M. Metody analizy danych marketingowych. Wydawnictwo Naukowe PWN. – Warszawa, 1996. – P. 121.

The conditions of criteria of the segmentation of customers are as follows: ([5], p. 48):

- they should make it possible to distinguish the segments so that each segment could have separate characteristics and could be operated by means of a specific marketing strategy,
- a distinguished segment must be big enough so that investing the time and effort into planning and exploiting it is justified,
- every distinguished segment must be easy to describe or determine so that it is possible to communicate with customers from this segment by means of a specific strategy of sales, promotion and advertising,
- every segment should have specific "purchase situation" - peculiar factors influencing the decision about the purchase or shaping the buyer's behaviour,
- the company should have the opportunity to change the structure, information systems and decisions so that it could be orientated to new segments.

A correctly conducted process of the segmentation should serve the realization of the above assumptions. According to M. McDonald and I. Dunbar it should include the following stages.

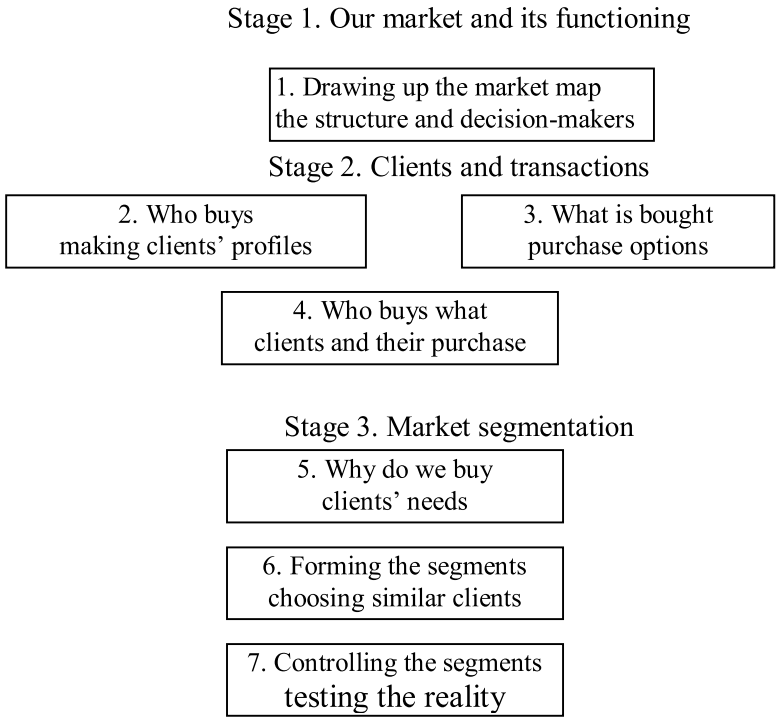

Fig. 1. Diagram of the process of segmentation

Source: McDonald M., Dunbar I.: Segmentacja rynku. Przebieg procesu i wykorzystanie wyników. Oficyna ekonomiczna, Kraków 2003, p. 50.

In the second stage of the research devoted to analysis of customers and their purchase, the authors recommend maximum possible simplification of the number of parameters with the help of the following procedures:

- eliminating obvious parameters – those which must meet elementary requirements so that the customer considers their purchase,
- concentration on groups of significant parameters,
- limiting the range of parameters (it concerns those values which are essential for customers and help diversify them),
- linking parameters and groups of parameters depending on the scheme of the purchase and categories accepted by customers,

- linking correlated parameters – those entailing values of other parameters or those appearing altogether.

In order to determine diversification of customers in terms of studied parameters the following statistical measures of variability are used: standard and average deviation, examining diversification of parameters in the examined group with regard to the average, and particularly the coefficient of variation that enables determining the strength of diversification of the parameters studied in the attempt.

As a rule a pilot attempt consisting of at least 100 observations is a basis of the segmentation process ([4]). Because of the need to widen the acquired results to the entire considered population, the methods of parameters' distribution which enable determining average characterizations of consumers in particular segments can as well be implemented in the process of segmentation. Statistical hypotheses verifying the credibility of received results of measurements with the assumed level of the statistical error belong to them, as well as credibility intervals determining the probability of reaching a particular level in the entire examined community by the examined parameters.

Methods of hypotheses verification - can be applied to confirm the statistical relevance of the studied parameters (level of the average, deviations), their possible correlation (examining the relevance of correlation coefficient), as well as in very distribution of consumers into segments – in order to assess the possibilities of combining two classes it is possible to conduct a test of the equality of their averages.

The elimination of correlated parameters is particularly associated with the notion of examining the correlational interdependence of parameters in statistics and econometrics, determined as the phenomenon of the catalysis or the collinearity. The *catalysis effect*, introduced and described by Z. Hellwig indicates strong correlation between explanatory variables[2] leading to the increase in value of the correlation coefficient, even when explanatory variables do not bring the information about the explained variable. According to T. Michalski *collinearity* appears when at least one of explanatory variables is strongly correlated with the remaining ones and Z_{nxk}[3] rank of matrix equals the number of considered explanatory K-variables. A difference between both phenomena results from the fact that after catalysis the strength of relations between explanatory variables is presented in comparison to the strength of their connection with the explained variables. Eliminating the effect of the catalysis results from formal accounts (improving the factual interpretation of the model and its parameters' relevance) and economical accounts (lowering the costs of data acquisition and model construction) which can constitute justification for considering it in selecting the criteria.

Lists received as a result of above mentioned operations include so-called "*principal distinctive parameters*". Establishing them requires conducting statistically correct *market research* and cannot depend on the subjective evaluation. Objectivity of the conducted evaluation can be assured by the correlational analysis and methods of the selection of explanatory variables, in which as a point of reference it is possible to accept the variable depicting the aim of introducing the segmentation e.g. amount of customers' expenses on sold products.

Correlation dependency (statistical) is a special variant of the stochastic relation appearing when it is possible to assign certain values of one variable to certain averages of several values of the second variable.

Correlation analysis consists in measuring the adherence between two or more variables, assuming that examined variables are of random character, and observations describing them are stochastically independent and come from the normal distribution. This approach does not deliver information about cause-effect relations of the variables, but it enables examination of the character and the level of variables' co-variability. Establishing the correlation of variables with the help of the correlation analysis and regression becomes peculiarly necessary when it is impossible to reveal the relation of two parameters in every individual case, but only in a numerous enough number of cases.

[2] Variables are those parameters, the values of which are statistically relevant in variability.

[3] The marking of a matrix of observation of potential explanatory variables. Its size is determined by the number of gathered observations (n) and the number of considered explanatory variables (K).

Methods of the selection of variables are very diversified and they are based on various criteria of the selection of explanatory variables. The most often used methods of the selection of variables are: Hellwig's method, defining the coefficient of multiple correlation, methods of analysis of the matrix of correlation coefficients, graphic methods and factor analysis. Methods of indicators of information capacity by Hellwig and defining multiple correlation coefficients are used to maximize a determined measure of the Q factor of the influence of explanatory variables on the explained variable. Graphic methods and the method of analysis of the matrix of correlation coefficients are conducted in successive sequences which serve to ensure right selection of variables:

- strong correlation of explanatory variables with the explained variable,
- weak correlation of explanatory variables.

Method of factor analysis enables transforming a big number of real indicators (variables) into a smaller number of conceptual variables. The fundamental assumption of this method is the possibility of isolating so-called *shared factors* or *principal components* connected with real variables, which enables real variables to be classified into homogeneous sub-families that are strongly correlated with individual factors. Usually there are a few shared factors or in special cases – one.

Factor analysis includes:

- *Principal Factor Analysis* (PFA),
- *Principal Component Analysis* (PCA).

This method can be treated both as the analysis method of the structure of the set of observations and as the analysis method of internal relations between variables. In the first application it can be useful in examining behaviours (reactions) of consumers on the market ([8]). Principal components analysis is more often applied in the selection of variables.

The core of Principal Component Analysis is transformation of the n set of the X real variables into a few variables Z_1 and Z_2 which are linear combinations of the X variables and moreover are not-correlated or normalised (the sum of squares of coefficients of the linear combination equals 1) and they approximate the vector of the X variables appropriately. In order to determine two principal components it will be sufficient to find values and correlation matrix's eigenvectors of the set of observations of primary variables[4]. Criterion of goodness of the approximation of a multidimensional observations' set is a minimization of the sum of squares of distances of points corresponding to multidimensional observations from points being their projections onto a plane which goes through the centroid of the set of observations[5] and is spanned on vectors of coefficients a_1 and a_2 of linear combination of equations of variables Z_1 and Z_2 from X ([2]).

All the aforementioned methods are based on Pearson's linear correlation coefficient. This measure characterizes the interval scale data, however in case of the data measured on so-called lower scale of measurement it is advisable to use other adequate coefficients, such as:

- for the nominal scale - based on statistics χ^2: coefficients of Czuprow, Cramer, Pearson, Youl or Hellwig. Because of the restrictions of some of the enumerated measures associated with the dependence on the size of the data presenting board, coefficients of Czuprow, Cramer and Hellwig are most often applied,
- for the ordinal scale – Spearman's rank coefficient or Kendall's coefficient.

A distribution of consumers into homogeneous groups is the most important component of the process of segmentation. K. Jajuga [2] enumerates the most popular methods of multidimensional analysis, used in the process of forming segments, that constitute the 6th step of the third stage of the segmentation specified in Fig. 1. Those are the methods of classification that enable distinguishing the market segments including "similar consumers". Those methods are:

- hierarchical methods,
- distribution methods:

[4] In order to eliminate the susceptibility of analysis results to units of measurement and orders of magnitude of the examined variables (member components are influenced by the variables of high values).

[5] It is the point corresponding to the vector of averages.

- area and density methods,
- optimization methods,
- graphic presentation methods.

In the *stochastic approach* the notion of *class*, with reference to methods of classification is understood as:

- set of observations of a unimodal distribution,
- set of observations of component distribution constituting a part of a mix of distributions,
- set of observations for which the value of density distribution function is bigger than a particular constant.

In a *descriptive approach* a *class* is understood as a set of observations characterized by a great similarity, that is a short distance. The distance is to determine the homogeneity of the class.

The most popular methods of classification are *agglomerative hierarchical methods*. They include the following stages:

1. A distance matrix of the set of observations (D) of *n*x*n* dimensions is formed, the elements (d_{ij}) of which are distances between observations of variables (x_i i x_j).
2. It is assumed that each observation constitutes one class.
3. A pair of classes, between which a distance is the smallest, is located.
4. The two classes are combined and the number of classes is reduced by 1.
5. A distance of the newly established class from other classes is determined.
6. Steps 3 - 5 are repeated *n*-1 times until all the classes are combined in one class.

Depending on the way of calculating the distance between two classes of combined elements the following methods are distinguished:

- *The nearest neighbour method* (*single linkage*) - the distance between classes is determined as the minimal distance between observations included in those classes.
- *The furthest neighbour method* (*complete linkage*) - the distance between classes is determined as the maximum distance between observations included in those classes.
- *The method of between-class average* - the distance between classes is calculated as the arithmetic average of the distances between observations included in those classes.
- *The median method* - the distance between classes is determined as the median of the distances between observations included in those classes.
- *The centroid method* - the distance between classes is determined as the minimal distance between the centroids (vectors of averages) of those classes.

Deglomerative hierarchical methods are applied according to the opposite principle:

1. A distance matrix of the set of observations (D) of *n*x*n* dimensions is formed, the elements (d_{ij}) of which are distances between observations of variables (x_i i x_j).
2. It is assumed that all observations constitute a class containing *n* number of elements.
3. At every stage of distribution a pair of classes, between which the distance is the largest, is located.
4. Classes containing those observations are divided in the following manner: two observations are assigned to different classes, and then the remaining observations of a divided class are assigned to them. The number of classes increases by 1.
5. A distance of the newly established class from other classes is determined.
6. Steps 3 - 5 are repeated *n*-1 times until *n* number of classes are combined (*n*-1 times).

The above description of applying hierarchical methods includes all sequences of the conducted classification (grouping or distribution), however, usually also in case of segmentation, the researcher is interested in one of them. For a determined number of *K*-classes it is recommended to stop the distribution after (*n*-*K*) stages (grouping) or (*K*-1) stages (distribution).

Determining the number of required stages of classification can be conducted on the basis of formal or factual criteria. K. Jajuga ([2]) argues that analysis of within-group and between-group variability of the values of parameters may be helpful in establishing the legitimacy of *r*-stage of classification.

The measure of the *between-group variability* is a sum of the distances of the centroids of groups of objects from the centroid of all examined objects, and the *within-group variability* is evaluated on the basis of the sum of the distances of objects in the class from the class centroid.

The process of grouping should be conducted until a considerable increase in within-class variability and decrease in between-class variability take place, which means combining two essentially different classes. In this situation the final stage of grouping should be cancelled and the procedure should end at the last but one stage.

In case of distribution the process should be cancelled when there is a considerable increase in between-class variability and a decrease in within-class variability. It means a correct distribution of differing observations, however their further distribution is not justified because of great between-class differences. The procedure finishes at this stage.

Specifying the "relevance" of increasing or decreasing in variability seems problematic. According to the author, a way to determine the appropriate number of classes may be determining the boundary value for the standardised distance between observations, that will guarantee their classification into more or less distant in terms of measured parameters. Classification proceeds until the shortest distance between classes of classes exceeds the boundary value.

In applying hierarchical methods the following graphical presentations are often used: dendrogram illustrating hierarchical process of grouping (distribution) and spanning tree, the vertices (observations) of which are combined if they are located in the nearest neighbourhood (have the shortest distances).

M. Walesiak enumerates the following advantages of hierarchical methods of grouping:

- they are based on one procedure,
- there is a possibility of supervision during the process of classification,
- it is possible to present results in the graphical form e.g. as a tree showing the order of connections between classes which enables determining mutual locations of classes and objects included in them,
- it is possible to conduct such methods with the help of many statistical programmes e.g. SPSS, Statistica, Systat, Statgraphics.

Distribution methods belong to the second group of classification methods. In order to eliminate so-called "chain effect" which is characteristic of hierarchical methods, two other methods have been elaborated, namely area and density methods.

Classes in *density methods* are created in the multidimensional space as areas characterised by a high density and separated from objects of smaller density. Those are for example graph method by Pluta and TAXMAP method by Carmichael and Sneath.

Method of spheres and method of full classes belong to *area methods*. In the method of spheres classes are star-shaped which means that they constitute a set of objects, the resemblance of which to the determined object (star centre) is not smaller than accepted threshold value. Full class is formed by objects, mutual resemblance of which is not smaller than the accepted threshold value.

Iterative optimization based on optimizing the function which characterises the quality of classification is applied for a large amount of observations. For some of these methods the optimization function may be a relation of a between-group diversity to a within-group diversity. From the start this method accepts a certain classification of objects, on the basis of e.g. hierarchical methods. The existing classes are then broadened by those observations which, after the transfer, increase (decrease) the values of optimization functions. Observations are transferred to the classes of the greatest increase. Methods of iterative optimization are as follows: J. P. MacQueen's method of *k*-averages, Wishart's method of *k*-averages, as well as Forgy's and Jancy's methods and ISODATA by Ball and Hall.

Methods of the graphical presentation of multidimensional observations are based on presenting the observations on the plane in order to enable visual analysis of the set of observations. Those methods reveal parameters of the examined phenomena and are based on a descriptive approach.

Methods created by Polish statisticians are often used in the process of classification. Those are e.g. Hellwig's method of linear organisation in which operations are arranged from "the worst" to "the best". It is possible to apply those methods only to variables that are stimulants, nominants or destimulants, the

interpretation of which allows observations to be organised lineary. They enable the transformation of multidimensional objects into the one-dimensional space.

The interpretation including determining distinctive parameters of created classes and differences between them is of high importance with reference to results of the classification. For data measured by means of interval or ratio scale, the results are interpreted on the basis of the determined centroids of classes[6] and standard deviations of variables in particular classes. In case of classification based on variables of nominal or ordinal scale, it is only possible to characterise particular classes for each variable in a descriptive (verbal) way.

M. Walesiak ([8]) enumerates a wide range of methods applied in the market segmentation. Apart from the aforementioned methods he distinguishes:

- conjoint measurement,
- automatic interaction detection (AID analysis),
- discriminant analysis,
- canonical correlation analysis (CCA) – to a limited degree.

Methods of *conjoint measurement* are aimed at determining the overall influence of two or more independent nominal variables on the variable measured by means of ordinal, interval or ratio scale. A result of applying those methods is a matrix of worth coefficients. The procedure of the method is as follows:

1. On the basis of distinguished characteristics and levels corresponding to them, a set of hypothetical products is formed. The number of products is a ratio of the number of levels corresponding to all characterisations of the products.

2. A set of respondents of the examination is selected. Respondents are asked to evaluate hypothetical products according to the tendency to their purchase.

3. Estimation is conducted with the help of appropriate methods e.g. regression analysis of part worth values of given levels of variables for a respondent, by means of creating the matrix of part worths:

- Matrix's lines are the subsequent respondents and its columns correspond to the number of distinguished levels for variables,
- Evaluation attributed by the respondent to products is a dependent variable,
- The influence of every level of the variable is taken into account by entering artificial variables into the model. The number of artificial variables must be smaller by one than the number of variants of the nominal variable.

Conjoined measurement method allows for distinguishing classes of consumers of similar preferences of the purchase which means that it can be applied in the process of segmentation. By estimating the total partworth of products for individual respondents and the whole community of respondents it is possible to determine their expected market share.

In *automatic interaction detection* (*AID*) there is one dependent variable measured on the interval or ratio scale, the variability of which is explained by a binary independent variables measured on the nominal scale to (before conversion primary variables can concern the nominal, ordinal, interval and ratio scale). The division of potential consumers into classes according to their preferences is aimed at distinguishing homogeneous groups of variability lower than in the overall statistical attempt. A measure of the variability is a sum of squares of deviations of empirical values of a variable dependent on the average value.

Algorithm of AID method is similar to methods of hierarchical grouping of variables and it is as follows:

1. All independent variables influencing a distinguished dependent variable are determined and recoded into binary variables.

2. For every two classes distinguished on the basis of a binary independent variable, a between-class and within-class sum squares of deviations is calculated.

[6] Arithmetic means of principal values constituting a class.

- Such a binary variable which has maximum value of between-class variance and minimal value of within-class variance is selected as the first one.

3. On the basis of variants of previously chosen binary variable, it is possible to divide the observations of a dependent variable into two classes.

4. Steps 2 and 3 are repeated until:

- distinguished classes contain too few observations (minimum 30),
- distinguished classes are relatively homogeneous (marginal within-class variability),
- there are no independent variables which considerably increase the relative homogeneity of classes.

One of the methods of *discriminant analysis*, namely the method of *Fisher discriminant coefficients* is mainly used in marketing research. In the process of segmentation this method enables to identify variables differentiating particular segments. A starting point is a dependent variable measured most often on the nominal scale and a set of independent variables measured on interval or ratio scale. Stages of conducting the method are as follows:

1. A set of examined consumers is initially divided into classes corresponding to variants of a dependent variable. Arithmetic means of the values of independent variables are calculated in individual classes – they constitute centroids of those classes.

2. The values of coefficients of the linear discriminant function are calculated on the basis of a criterion of the maximization of the function of a variance ratio of between-class variance and within-class variance.

Values of the discriminant function are a base of deciding on the membership of examined objects in new classes. Moreover, on the basis of coefficients it is possible to estimate to what degree independent variables influence between-class differences. Discriminant analysis shows certain resemblance to methods of the *k*-averages optimization, however, in its case it is obvious from the beginning how many groups there are and where each object belongs to, while in averages analysis there is no such information.

Canonical correlation analysis is applied in order to examine relations between two groups of variables which are isolated in a natural way. A base of inferences may be both a correlation degree of particular variables in groups with a canonical variable and the values coefficients standing by the variables in linear functions of canonical variables. *Canonical variables* are normalised[7] linear combinations of variables included in selected groups. In this method the values of equations' coefficients are determined in such a way to ensure maximum value of a linear correlation coefficient between two canonical variables. In practice a large number of canonical variables is possible, however, the first pair having the biggest coefficient of linear correlation is the most important. This method is rarely applied in market research because of the difficulties in results interpretation.

Conclusions. Quantitative methods are often applied in the process of customers' segmentation. They can be applied at all stages of distribution of customers both by using special measures (average, variance, variability coefficient), estimators (methods of verifications of hypotheses), as well as specific algorithms. Hierarchical methods are the most popular among applied methods of classification. Their standardised procedure and frequent applications caused a number of computer tools to appear. Certain limitation of this group of methods is defining the criterion of optimal distribution of classes. Ambivalent notion of "relevant increase / decrease in variability" concerning the moment of finishing the procedure enables other methods of distribution to be applied e.g. iterative optimization.

Usefulnesses of applications of described methods is proved by their frequent practical implementations, however a possibility of applying determined measures or calculations only for selected scales of the measurement is their major limitation. Marketing research in the scope of segmentation is based on a variety of criteria. Used data can concern properties of the nominal scale (sex, place of residence), ordinal scale (preferences of the purchase) as well as higher scales (age, affluence, etc.). Diversity of applied scales constitutes a major limitation of the possibility of applying the aforementioned

[7] Variance equals one.

methods. Method of canonical correlation and the interaction detection analysis require variables measured on the interval, ratio and nominal scales (binary variables). Conducting factor analysis is possible only for data measured on interval or higher scale with the assumption that variables are subject to a multidimensional normal distribution. Transferring the data measured on so-called "higher scales" (interval, ratio scales) into the data from lower scales (nominal, ordinal scales) is possible to conduct by means of transformation ([6], [8]) with the method of parameters' standardisation or the method of ranks, however it is connected with a considerable loss of information and low accuracy of the results of the method. Transfer from lower to higher scales is impossible.

Prospects of the future research. The article presents theoretical analysis of the scope of quantitative methods which can be used in the process of segmentation. Additional information concerning the usefullness of those methods, as well as their virtues and limitations, can be obtained after their practical implementation and distinguishing segments of the examined group of purchasers on the Polish market.

Further research will be devoted to characteristics of students of privately owned colleges, their economic situation and preferences concerning their educational specialisations. Those properties will be analysed with regard to a group homogeneity and segments possible to distinguish.

1.Gatnar E., Walesiak M.(eds.).: Metody statystycznej analizy wielowymiarowej w badaniach marketingowych, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 2004. 2. Grabiński T., Wydymus S., Zeliaś A.: Metody doboru zmiennych w modelach ekonometrycznych, PWN, Warszawa 1982. 3. Jajuga K.: Statystyczna analiza wielowymiarowa. Wydawnictwo naukowe PWN. Warszawa 1993. 4. Kotler Ph.: Marketing od A do Z. PWE. Warszawa 2004. 5. Kotler Ph., Armstrong G., Saunders J., Wong V.: Marketing. Podręcznik europejski. PWE, Warszawa 2002. 6. McDonald M., Dunbar I.: Segmentacja rynku. Przebieg procesu i wykorzystanie wyników. Oficyna ekonomiczna, Kraków 2003. 7. Mruk H.(ed.): Analiza rynku. PWE. Warszawa 2003. 8.Walesiak M.: Metody analizy danych marketingowych. Wydawnictwo Naukowe PWN, Warszawa 1996.