DISCRIMINANT ANALYSIS OF OFFICE PROPERTY OBJECTS DISTRIBUTION ACCORDING TO THE PRICE LINES (exemplified by Samara)

© 2009 E.V. Zarova, S.P. Krivozubov*

Keywords: discriminant analysis, statistic modeling, office property, administrative district, local market, pricing.

The present article provides the theoretical grounds and methods of the practical realization of the discriminant analysis pursuing the aim to create an adequate system of the pricing static modeling of the office property objects.

In practice, when it comes to the real property appraisal one of the most difficult-to-solve and important problems is the problem of objectiveness and adequateness of choosing the model objects in order to define the price by comparative method for the real property objects that would be similar to them.

As the researches showed, done by the authors on the basis of over 1400 office objects overall data that were exposed and sold at the real estate market of Samara in 2007, in order to solve this problem one needs to apply the set of methods of the multifactor statistic analysis that enables to find out the most typical groups of objects not only on the basis of their observed and measured characteristics but also taking into consideration the interaction of their factor features which produce a complex influence upon the price of the object.

Solution of the problem of how to pick out the groups with similar office property objects based on the price level of realization for 1 sq. m. (V15), while each object is characterized by the system of factor features, is only possible by means of applying the procedures of the discriminant analysis. The last two groups of methods include: 1 - methods of observation classification by groups; 2 - methods of interpretation of inert-group differences. The methods of classification on the basis of the set of features make it possible to put an object under observation into one groups. The methods of interpretation of the inter-group differences make it possible to find out whether it is realizable to differentiate one class from another with the help of the present set of variables as well as to identify the most informative features which precondition the structure of the groups that had been picked out.

A remarkable characteristic of the discriminant analysis as a method of classification is the fact that on the basis of the preliminary analysis of the input data the number of groups (classes) that will divide the set of objects under observation is known beforehand. The aim now is to construct the biding rule that would enable, on the basis of the objects parameters measurements, to mark the group it relates to. The number of groups is available beforehand like the fact that the object relate to certain group initially. Such a method of classification is also called the method of pattern recognition "with guidance of a teacher".

On the basis of the structural group according to the figure V15, one make a conclusion a priori that the office property objects according to the price level can be divided into our classes (price: "very low", "low", "high", "very high") (table 1).

In accordance with the methods of the discriminant analysis, there comes the formation of guiding selections on the basis of which it is possible to make a classification of both the objects that did not comprised the guiding selections and all other objects that should be divided. The calculations were done with the help of the Stepwise Discriminant Function "Statistica 6.0".

At the initial stage of the analysis of the guiding selections formations, they are checked for adequateness on the basis of the statistical criteria: generalized distance and a posteriori probability of reference to the "intended" group.

In order to check and make sure of the adequateness of the object segmentation and distribution into classes according to the selling

* Elena V. Zarova, Doctor of Economics, Professor, Head of Statistics Department, Samara State University of Economics; Sergey P. Krivozubov, post-graduate student, Samara State University of Economics.

Table 1

Grouping of the office property objects according to the price of selling, rub. per sq. m (V15)

Price of selling, rub. per sq. m	Number of objects	Specific weight, %	Class of the price group
under 20000	99	7,02	Very low
20000-40000	853	60,45	Low
40000-60000	411	29,13	High
60000 and over	48	3,40	Very high
Total	**1411**	**100**	-

Table 2

Classification matrix at the first stage of the discriminant analysis of the office property objects complex (based on the results of the a priori distribution into four price classes according to the figure V 15)

Group Price class	Classification matrix (disc. analysis-2.sta) Rows: Observed classifications Columns: Predicted classifications				
	Persent Correct	high p=,29202	low p=,60541	Very high p=,03419	Very low p=,06838
High price	41,70732	171	225	14	0
Low price	86,94118	90	739	0	21
Very high price	18,75000	20	19	9	0
Very low price	43,75000	5	49	0	42
Total	**68,44730**	**286**	**1032**	**23**	**63**

price one should analyze the classification matrix where the columns show the afore-predicted observations together with the percent correct at each group and the overall percent correct at the whole selection (table 2).

As it appears out of the matrix, at the initial distribution of objects into the price classes for the first group ("high price") 171 objects were included correctly as opposed to 239 objects which were included incorrectly, while 225 of them should be included to the price group "Low price" and 14 - to the price group "very high price". One can arrive at such conclusions also concerning other price classes which had been selected on the basis of the initial structural grouping of the office property objects according to their selling price for square meter (V 15).

Looking at the classification matrix one can also make a conclusion that some objects were initially referred to the selected price classes correctly; though, there are objects that should not be in certain classes according to the statistical criteria. This is supported buy the coefficient of correctness for each group that is not equal to 100%, consequently, the overall coefficient of correctness is not equal to 100% (68,45%) (see table 2).

Insufficient level of adequateness of the initial classification of the office property objects into the class groups can be exemplified and illustrated by the Wilks lambda level the value of which under good discrimination should be as much close to zero as possible while the value of the F-criterion - as much high as possible (fig. 1). The Wilks lambda statistics (λ) is calculated as the ratio of the intragroup dispersions (covariance) matrix determinant to the covariance matrix determinant. The value of belong to the interval [0,1].

In order to solve the problem of how to get the correct guiding selections, it was reasonable to exclude from the initial (and further) guiding selections the objects that form a homogeneous group. For this very purpose, from the figure there were excluded the cases that had been referred to a certain group by mistake. Reference of objects to a certain group is considered to be an incorrect if the generalized distance from an object to the center of its group is much higher that the one from it to the center of other groups, while the a posteriori probability of reference to this very group is much lower than the critical value.

During the removal of objects from the group we also took into consideration the fact that

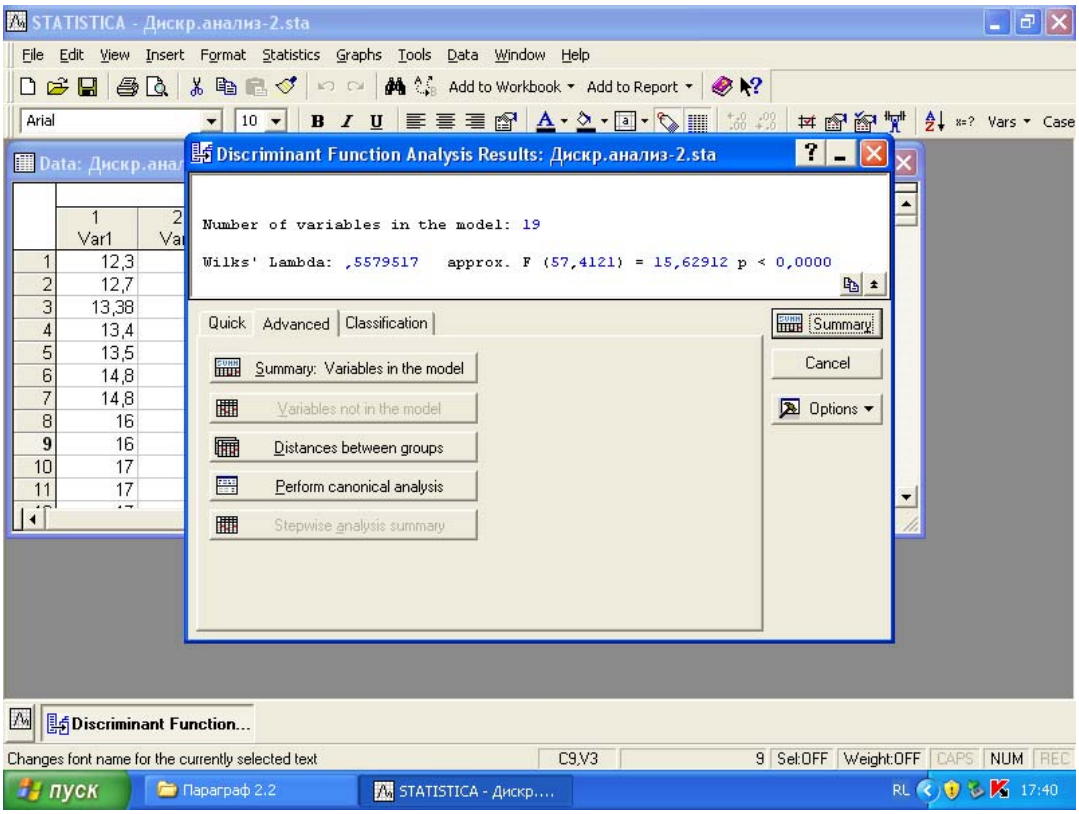

Fig. 1. **The results of the discriminant functions analysis at examining the initial distribution of the office property objects in Samara by the price groups**

with removal there is a shift of the gravity center (medium vector) since it is determined by the further observations.

The procedure of removal of observations is presented in the way of a scheme in fig. 2. It required eight consecutive transitions (steps) and lasted until the overall coefficient of correctness in the classification matrix reached 100%, while the value of the Wilks statistics was at the same time 0,0826, which after all, means a good discrimination of the final guiding selection (fig. 3).

Availability of variables in the models was estimated according to the level of tolerance. The value of tolerance was designed as 1 minus R-square, where R-square - the coefficient of the multiple correlation for the respective variable with all other variables in the present model. It corresponds to the part of dispersion that relates to the respective variable.

In accordance with the final classification matrix, the overall number of the objects that appeared to be in the guiding selection after all the objects, incorrectly referred to the classification groups, had been removed reached 867 (61,45% from the initial volume of the totality). Theses objects are to be distributed in the following way (table 3).

The guiding selection, received at the final step (the 8th one, fig. 2), make it possible to make a classification of the objects which had not been included into the final discrimination (these are 554 objects, or 38,55%) or calculate the classification values for the objects that are again to be exhibited at the market.

Table 3

Distribution of the office property objects from the final guiding selection by the group classes

Price class (by the price level, rub. for sq. м)	Number of objects	Specific weight in the selection, %
Very low price	34	4,00
Low price	704	81,02
High price	125	14,55
Very high price	4	0,005
Total	**867**	**100,00**

For this very purpose, the final value of the classification functions (S_i) is to be calculated by the following formula:

$$S_i = C_i + W_{i1} \cdot X_1 + W_{i2} \cdot X_2 + ... + W_{im} \cdot X_m ,$$

where i - the number of the totality (price group); j = 1, 2, ... m - the numbers of variables; C_i - the constant for the i- totality (i-classification function); W_{ij} - weights for the j-variable at calculation the classification exponent; For the i-totality (price group); X_i - the value of the factor variable j for the corresponding office property object.

The new object should be referred to the class (group) which classification value (S_i) is highest

The final view of the classification functions is represented by the following models (table 4).

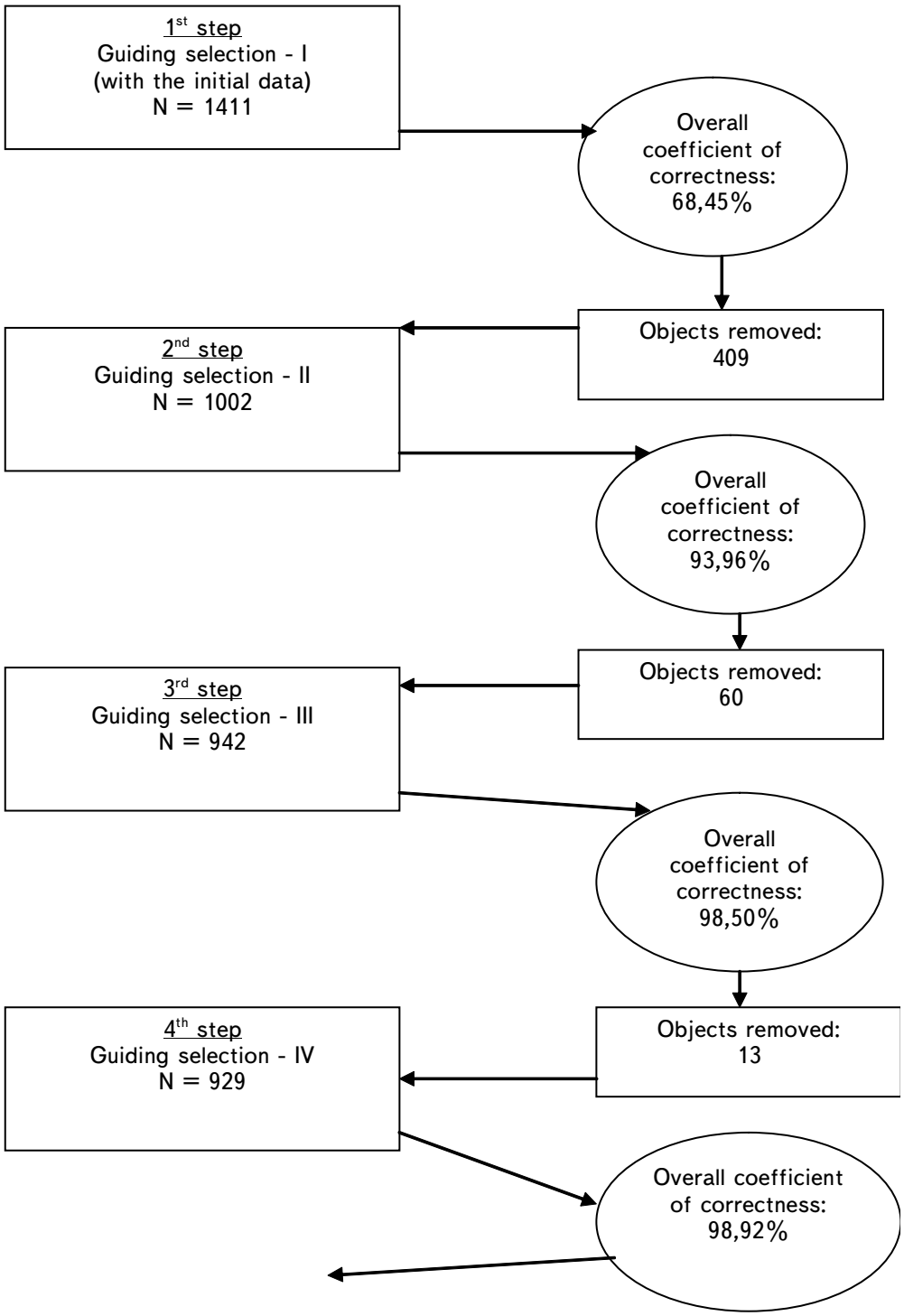

Fig. 2. **The algorithm of the discriminant analysis with the consecutive removal of the office property objects from the initial totality**

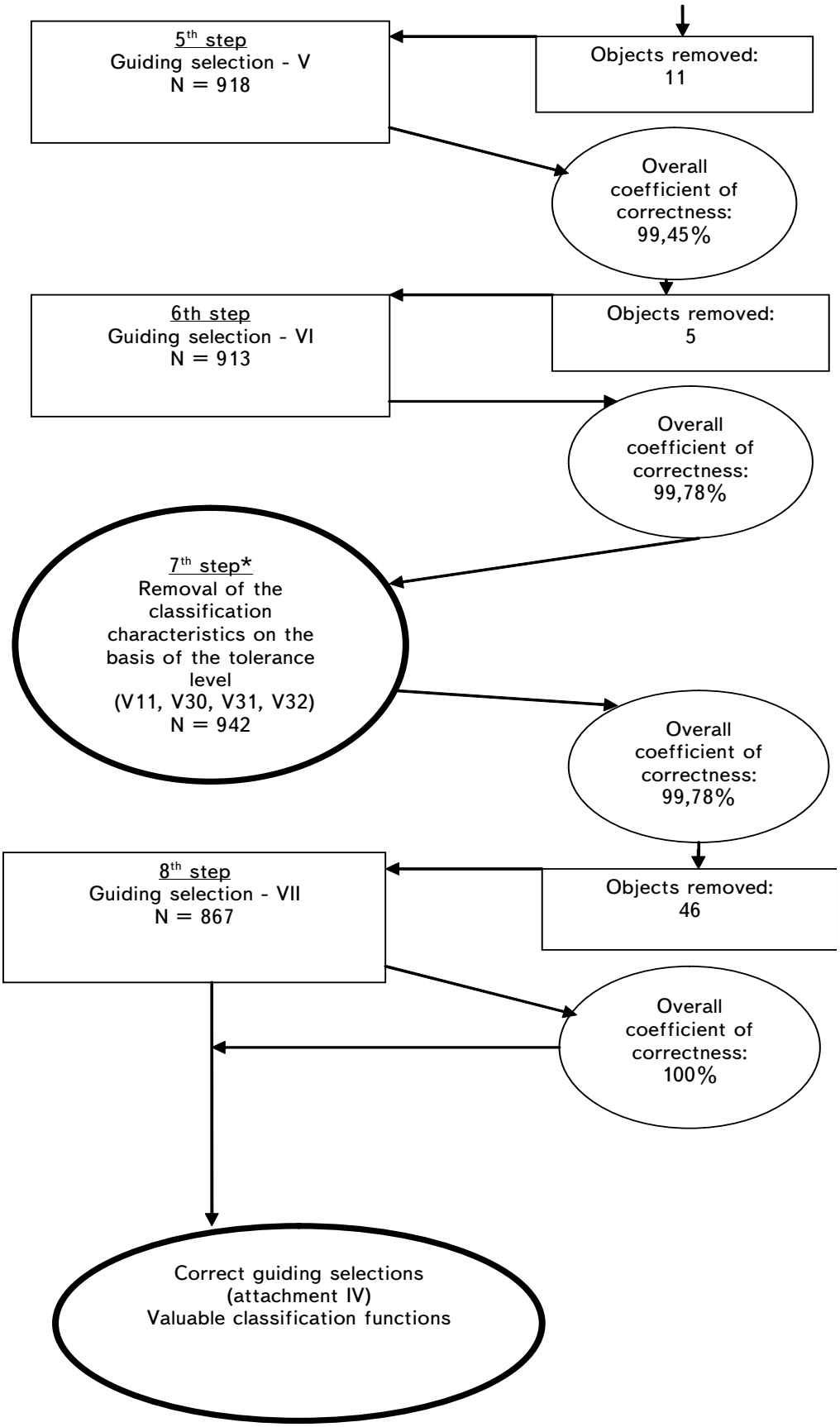

Fig. 2. **(the final part) The algorithm of the discriminant analysis with the consecutive removal of the office property objects from the initial totality**

* The variables had been removed because their tolerance level appeared to be lower than the critical one (0,01).

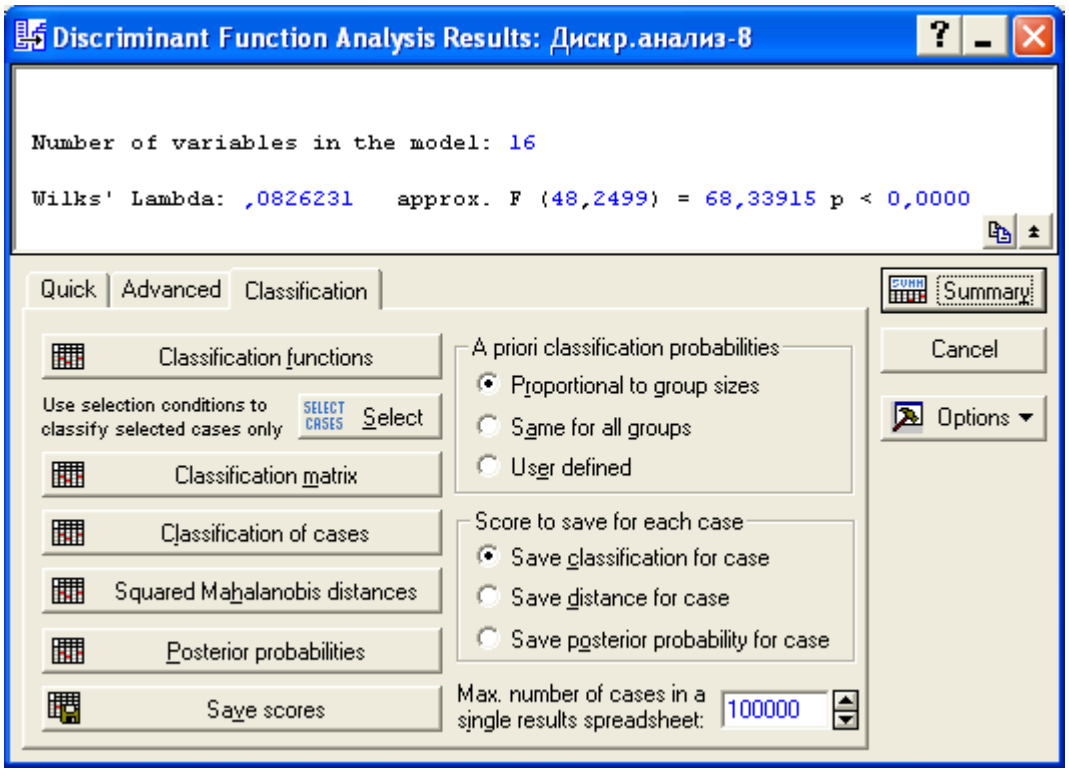

Fig. 3. **The results of the discriminant functions analysis at the final step (the 8th one) of the guiding selection formation of the office property objects distribution in Samara by the price groups**

On the basis of the received guiding selections there was done a recurring classification of the office property objects which had not been included in the guiding selections. As a result of their regrouping and in accordance with the calculated classification values (S_i), we managed to get the final grouping (fig. 4). The objects that formed after the discriminant analysis each of the groups have mutual patterns of the pricing influence from the factors under ob-

Table 4

Parameters of the classification functions according to the results of the discriminant analysis of the office property objects totality distribution in Samara by the price groups

	Parameters of the classification functions by the price groups			
	Low p=,81024	High p=,14552	Very low p=,03958	Very high p=,00466
Var1	0,0017	0,001	0,0138	0,004
Var2	0,9681	1,569	1,0097	2,018
Var3	0,9126	0,888	0,2810	1,028
Var4	0,0117	0,006	0,0614	0,006
Var22	26,4611	20,733	23,7545	27,216
Var23	0,0550	0,043	0,0493	0,057
Var24	0,0601	0,592	0,1382	0,218
Var25	0,0092	0,011	0,0074	0,026
Var26	0,0025	0,003	0,0022	0,004
Var27	-0,0732	-0,048	-0,2183	-0,643
Var28	-0,0004	-0,001	0,0004	0,002
Var29	2,8714	4,851	-0,3931	4,687
Var33	-0,0692	-0,160	-0,0478	-0,190
Var34	-0,1795	0,947	-0,2159	-0,160
Var35	4,1944	17,982	1,8117	7,677
Var36	0,0020	0,003	0,0019	0,003
Constant	-69,2762	-104,973	-74,6900	-140,635

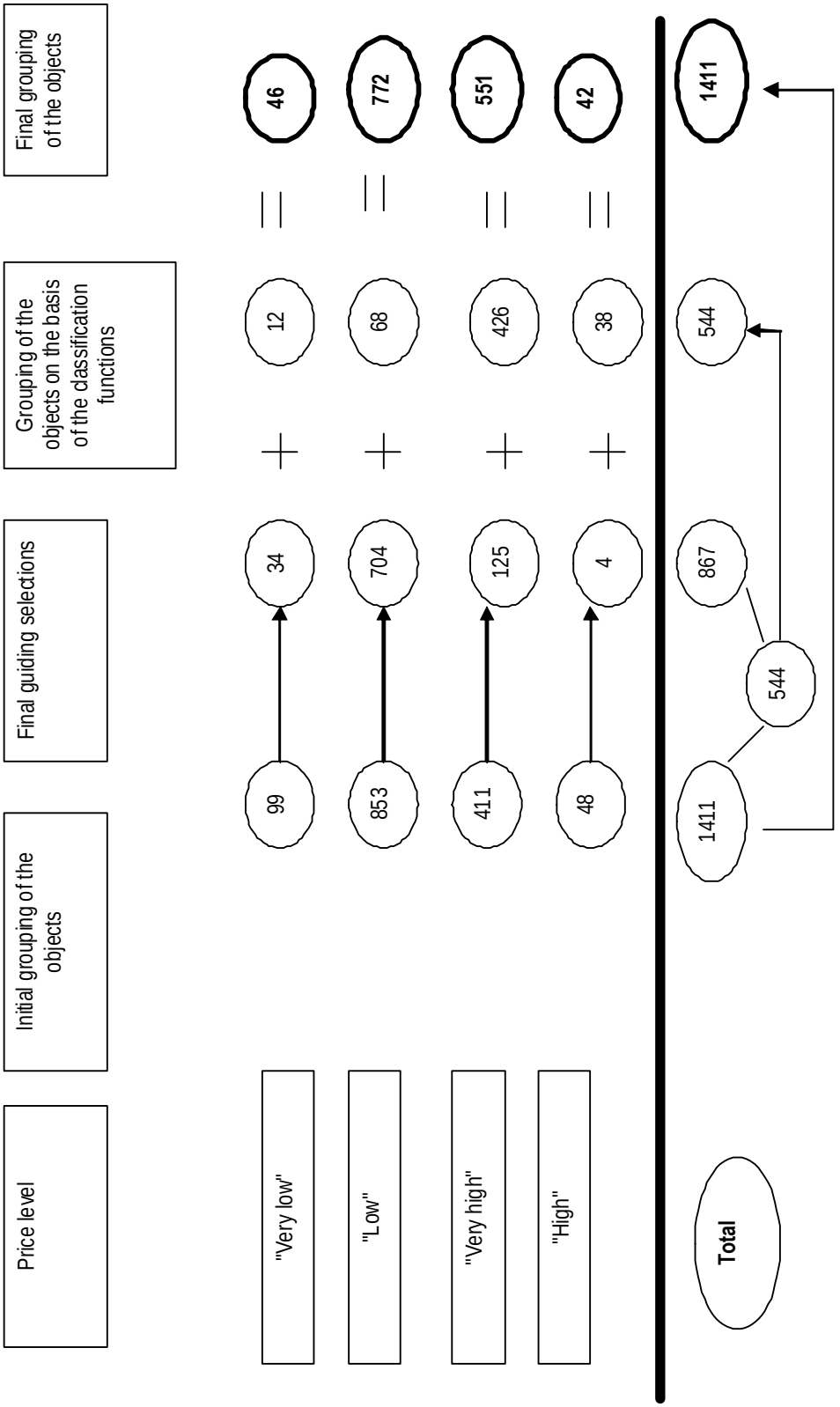

Fig. 4. **The scheme of the discriminant analysis procedure of the office property objects distribution in Samara by the price groups**

Table 5

The matrix of the dual coefficients of the office property objects correlation in Samara for 1 sq. m. (V15) and of the factor variables which caused the objects distribution by the price groups

Variable and its indication		Groups of objects by the price level			
		Very low	Low	High	Very high
Area, sq. m.	- V 1	0,36*	-0,02	-0,17	-0,18
Storey	- V 2	-0,27	-0,12	-0,02	-0,33*
Number of storeys	- V 3	0,51*	0,04	0,05	0,14
Exposition, days	- V 4	-0,35*	-0,07	-0,06	-0,14
Stop not far away (yes/ no)	- V 22	0,20	0,10	-0,01	0,05
Distance to the nearest stop, m	- V 23	0,21	-0,02	0,08	-0,19
Distance to the nearest street, m	- V 24	-0,20	-0,07	-0,06	-0,13
Distance to the nearest main highway, m	- V 25	0,33*	-0,03	0,13	-0,13
Intensity of the traffic, hours	- V 26	0,56*	0,10	0,01	-0,10
Route of the public transport	- V 27	0,85*	0,11	0,04	-0,19
Runs of the public transport a day	- V 28	0,80*	0,12	0,04	-0,13
Density of the cadastral quarter	- V 29	0,01	0,01	0,04	0,32*
Business activity, no	- V 33	0,50*	0,03	0,08	-0,02
Historical sites	- V 34	0,39*	-0,05	0,12	0,08
Historical site or not (yes/no)	- V 35	0,02	-0,05	-0,00	0,01
Price of the real estate, rub. for 1 sq. m.	- V 36	0,40*	0,01	0,25	0,42*

Note: * shading marks the dual coefficients of the correlation that are valuable according to the t-criterion of Student.

servation, which all combined and together with the force of influence provide reference of the objects to this or that price group. In order to explore these patterns we have calculated the dual coefficients of the office property object price correlation for 1 sq.m. (V15) as well as the factor variables that preconditioned the discrimination of the objects in accordance with the price groups (table 5).

As the table 5 demonstrates, the grouping influence of the factor variables has a valuable force, statistically, of the pricing influence only in two groups: 1 - "very low" price (variables V1, V3, V4, V25, V26, V27, V28, V33, V34, V36) and 2 - "Very high" price (variables V2, V29 and V36). The other groups of objects (with the price level "low" and "high") do not bear any statistically valuable influence of the grouping factor variables on the price level of the object for 1 sq. m.

The analysis of the factor variables influence pursuing the aim to eliminate the multicollinear links based on the correlation dual coefficients matrix made it possible to receive a structure of the object price regression models (V15) by the two groups: "very low" and "very high" prices. As a result of their parameters estimation and analysis, we managed to get the following models and patterns:

1. Price group: "Very low" price.

$$V15 = 20117{,}25 - 1{,}55\ V1 - 13{,}08\ V4 + 3{,}89\ V26 + 882{,}98\ V27 + 6544{,}59\ V34,$$

where V 15 - final price, rub. 1 sq. m; V 1 - area, sq. m; V 4 - exposition, days; V 26 - Intensity of the traffic, hours; V 27 - Route of the public transport; V 34 - Historical sites.

Table 6

Parameters and characteristics of the multi-factor regression pattern of the selling price level of the office property objects, rub. For 1 sq. m. (V15) by the price group "Very low price"

N=45	Results of the regression for the independent variable: Var 15 (Discr.-ready-very low) R= ,94644293 R2= ,89575423 Corrected. R2= ,88238938 F(5,39)=67,023 r					
	BETA	Стд.Ош.	B	Стд.Ош.	t(39)	p-level.
Absolute term			20117,25	1488,103	13,51872	0,000000
Var1	-0,836982	0,155908	-1,55	0,289	-5,36843	0,000004
Var4	-0,173478	0,063858	-13,08	4,814	-2,71664	0,009782
Var26	0,369829	0,087191	3,89	0,917	4,24160	0,000132
Var27	0,589412	0,076472	882,98	114,561	7,70758	0,000000
Var34	1,085975	0,163587	6544,59	985,851	6,63852	0,000000

Table 7

Parameters and characteristics of the multi-factor regression pattern of the selling price level of the office property objects, rub. For 1 sq. m. (V15) by the price group "Very high price"

N=42	Results of the regression for the independent variable: Var 15 (Discr.-ready-very high) R= ,56426011 R2= ,31838947 Corrected. R2= ,28343509 F(2,39)=9,1087 r					
	BETA	Стд.Ош.	B	Стд.Ош.	t(39)	p-level.
Absolute term			-15189,9	16116,23	-0,942524	0,351727
Var29	0,381757	0,133393	31124,2	10875,38	2,861895	0,006739
Var36	0,469536	0,133393	1,2	0,33	3,519941	0,001115

Statistical characteristics of the derived equity that confirms its adequateness and the value of the parameters are listed below in the table 6.

Comparison the regression standardized coefficients values: negative influence on the objects selling price level in the group under observation is produced by the factor variables: the greatest one - area, sq. m. (V1) and relatively less - the period of exposition at the open market, days (V4); positive influence is produced by: the greatest one - proximity of historical sites (V34), the average one - availability of the public transport (V27), the smallest one - intensity of the traffic (V26).

One can obviously arrive at a conclusion that the objects price level which can be referred to the classification group "the cheapest" (based on the price for 1 sq. m.) bears a direct dependence upon the development and prestige of the district, the area location as well as upon the demand level and, consequently, a diverse dependence upon the figures of competitiveness. For the heap objects the latter is as much lower as their area is larger, and its decrease influences the increase of the objects period of exposition of this very price segment at the open market.

2. Price group: "Very high" price.

$$V15 = -15189,9 + 31124,2\ V29 + 1,2\ V36;$$

where V 15 - final price, rub. 1 sq. m; V 29 - Density of the cadastral quarter; V 36 - Price of the real estate, rub. for 1 sq. m.

Statistical characteristics of the pattern are listed in the table 7.

As the figure 6 demonstrate, the approximating characteristics of the received pattern are far from to be high: the coefficient of the multiple determination $R^2 = 0,318$, in other words, the pattern provides 31,8% of the explained variation.

Comparison of the β (beta)-coefficients values proves that the price for 1 sq. m. of the most expensive objects of the office property is highly dependent upon the price for 1 sq. m. of the residential property objects (V36) in the corresponding district and also, to a lesser degree, still in the same way, upon the density of the cadastral quarter (V29) in which this district is located.

Thus, the price of the most expensive objects depends upon competitiveness at the local real estate market, micro-level on the city scale and overall price level of this market.

Unimportance of the correlation links in the two other price groups of the totality of the office objects under present observation (price: "low" and "high") together with the necessity to intensify the researches of the pricing influence of the factors leads to the necessity of transition to the next stage of the discriminant analysis - canonical analysis.

The methods, suggested as a complex, allows developing an adequate statistical panel in order to create multi-factor regression patterns of pricing for the office property objects in the city taking as a basis the conditions of the substantial homogeneity and the factor of the pricing.

1. *T.A. Dubrova, A.G. Bazhin, L.P.* Bakumenko Discriminant analysis in the system of "STATISTICA"/Work book. - M.: Moscow state University of economics, statistics and computer technologies, 2000. P. 57.

2. *A.A. Khalafyan.* STATISTICA 6 Statistical analysis of data, 3rd edition. textbook . - M.: LLC "Binom-Press", 2008. P. 512.