

## ДИСКРИМИНАНТНЫЙ АНАЛИЗ РАСПРЕДЕЛЕНИЯ ОБЪЕКТОВ ОФИСНОЙ НЕДВИЖИМОСТИ ПО ЦЕНОВЫМ ГРУППАМ (НА ПРИМЕРЕ Г. САМАРЫ)

© 2009 Е.В. Зарова, С.П. Кривоzubов\*

**Ключевые слова:** дискриминантный анализ, статистическое моделирование, офисная недвижимость, административный район, локальный рынок, ценообразование.

Приводятся теоретическое обоснование и методы практической реализации дискриминантного анализа с целью выработки адекватной панели для статистического моделирования ценообразования на объекты офисной недвижимости.

В практической деятельности по оценке недвижимости одним из основных и наиболее трудно разрешаемых вопросов является вопрос об объективности и адекватности установления объектов-аналогов с целью определения сравнительным методом цены для однородных по отношению к ним объектов недвижимого имущества.

Как показали исследования, выполненные авторами на основе данных более чем о 1400 офисных объектов, экспонированных и реализованных на рынке недвижимости г. Самары в 2007 г., решение этого вопроса требует применения комплекса методов многофакторного статистического анализа, позволяющего выявлять типичные группы объектов не только на основе их наблюдаемых и измеримых характеристик, но и с учетом взаимодействия их факторных признаков, комплексно влияющих на цену объекта.

Решение задачи выделения однородных групп объектов офисной недвижимости по уровню цены реализации за 1 кв. м (V15) при том, что каждый объект характеризуется системой факторных признаков, возможно путем применения процедур дискриминантного анализа. Последние включа-

ют в себя две группы методов: 1 - методы классификации наблюдений по группам; 2 - методы интерпретации межгрупповых различий. Методы классификации на основе набора признаков обеспечивают отнесение исследуемого объекта к одной из нескольких групп. Методы интерпретации межгрупповых различий позволяют выяснить, можно ли с помощью данного набора переменных отличить один класс от другого, позволяют установить наиболее информативные признаки, обуславливающие состав выделенных групп<sup>1</sup>.

Отличительным свойством дискриминантного анализа как метода классификации является то, что на основе предварительного анализа исходных данных заранее известно число групп (классов), на которые следует разбить рассматриваемую совокупность объектов. Задача состоит в том, чтобы построить решающее правило, позволяющее по результатам измерений параметров объекта указать группу, к которой он принадлежит. Число групп заранее известно, также известно, что объект принадлежит к определенной группе. Такой метод классификации еще называют методом распознавания образов "с учителем"<sup>2</sup>.

Таблица 1

**Группировка объектов офисной недвижимости  
по уровню цены реализации, руб. за 1 кв. м (V15)**

Цена реализации	Число объектов	Удельный вес, %	Класс ценовой группы
До 20000	99	7,02	Очень низкая цена
20000-40000	853	60,45	Низкая цена
40000-60000	411	29,13	Высокая цена
60000 и выше	48	3,40	Очень высокая цена
<b>Итого</b>	<b>1411</b>	<b>100</b>	-

\* Зарова Елена Викторовна, доктор экономических наук, профессор, зав. кафедрой статистики Самарского государственного экономического университета; Кривоzubов Сергей Павлович, соискатель, Самарский государственный экономический университет.

**Классификационная матрица на первом шаге дискриминантного анализа совокупности объектов офисной недвижимости (по результатам априорного распределения на четыре ценовых класса по показателю V 15)**

На основании структурной группировки по показателю V15 можно сделать априорное заключение о том, что объекты офисной недвижимости по уровню цены могут быть разделены на четыре класса: “Очень низкая цена”, “Низкая цена”, “Высокая цена”, “Очень высокая цена” (табл. 1).

В соответствии с методикой дискриминантного анализа формируются обучающие выборки, на основе которых возможно произвести классификацию как объектов, не вошедших в обучающие выборки, так и любых других объектов, подлежащих разбиению.

и 14 - в ценовую группу “Очень высокая цена”. Подобные заключения можно сделать и по остальным ценовым классам, выделенным на основе исходной структурной группировки объектов офисной недвижимости по цене их реализации за 1 кв. м (V15).

По классификационной матрице можно сделать вывод, что часть объектов была правильно изначально отнесена к выделенным ценовым классам, но есть объекты, которые по статистическим критериям не должны находиться в определенных классах. Этому соответствует коэффициент корректности по

Расчеты выполнялись с помощью модуля Stepwise Discriminant Function (Дискр. анализ 2 шаг) в программе SPSS “Statistica 6.0”. Анализ модуля дискриминантных функций показал, что на первом этапе анализа по одной группе объектов офисной недвижимости формирования обучающих выборок они проверяются на корректность на основе статистических критериев расстояния Махаланобиса и априорной вероятности попадания в “свою” группу.

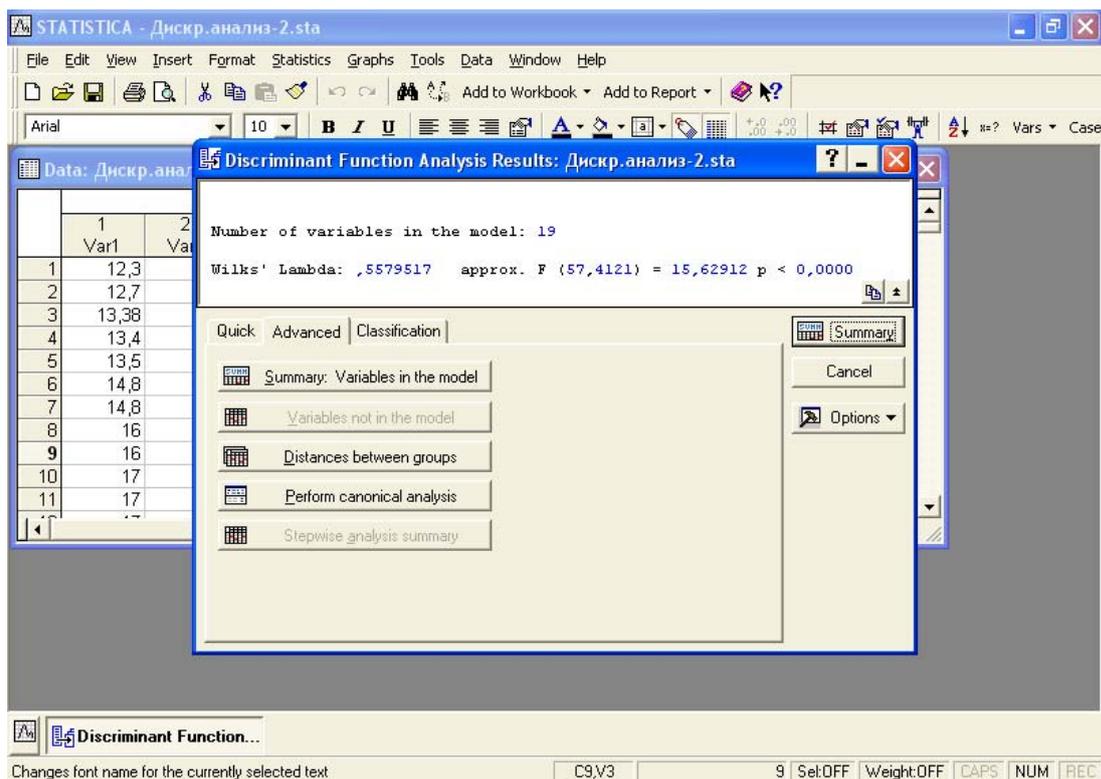
Ценовой класс	Высокая цена	Низкая цена	Очень высокая цена	Очень низкая цена
Высокая цена	170732	225	14	0
Низкая цена	8694118	739	0	28
Очень высокая цена	1875000	20	19	0
Очень низкая цена	4375000	5	49	42
<b>Итого</b>	<b>6844730</b>	<b>286</b>	<b>1032</b>	<b>63</b>

Для проверки правильности априорного распределения объектов на классы по уровню цены реализации следует проанализировать классификационную матрицу (Classification matrix), в которой столбцы показывают предсказанные наблюдения, а также процент корректности (Percent Correct) по каждой группе и общий процент корректности по всей выборке (табл. 2).

Из матрицы следует, что при первоначальном распределении объектов на ценовые классы к первой группе (“Высокая цена”) правильно были отнесены 171 объект и неправильно 239 объектов, из которых 225 следовало отнести в ценовую группу “Низкая цена”

и 14 - в ценовую группу “Очень высокая цена”. Подобные заключения можно сделать и по остальным ценовым классам, выделенным на основе исходной структурной группировки объектов офисной недвижимости по цене их реализации за 1 кв. м (V15). По классификационной матрице можно сделать вывод, что часть объектов была правильно изначально отнесена к выделенным ценовым классам, но есть объекты, которые по статистическим критериям не должны находиться в определенных классах. Этому соответствует коэффициент корректности по расчету, который не равен 100%, тогда как по шагам не равен 100% и общий коэффициент корректности (68,45%) (см. табл. 2). Недостаточный уровень адекватности исходных данных по классификации объектов офисной недвижимости по ценовым группам подтверждается уровнем лямбды Уилкса (Wilks lambda), значение которой при хорошей дискриминации должно быть как можно ближе к нулю, а значение F критерия как можно больше (рис. 1). Статистика лямбды Уилкса ( $\lambda$ ) вычисляется как отношение детерминанта матрицы внутригрупповых дисперсий (ковариаций) к детерминанту ковариационной матрицы. Значения  $\lambda$  принадлежит интервалу [0,1]<sup>4</sup>.

Для решения задачи получения корректных обучающих выборок следовало исключить из начальной (и последующих) обучающих выборок те объекты, которые по своим показателям не соответствуют большинству объектов, образующих однородную группу. С этой целью из таблицы классификации объектов исключались случаи, некорректно отнесенные к определенной группе. Отнесе-



**Рис. 1. Результаты анализа дискриминантных функций в исследовании исходного распределения объектов офисной недвижимости в г. Самаре по ценовым группам**

ние объектов в конкретную группу считается ошибочным, если расстояние Махаланобиса от объекта до центра его группы значительно выше, чем от него до центра других групп, а апостериорная вероятность попадания в эту группу ниже критического значения.

При удалении объектов из группы учитывалось также, что при удалении смещается центр тяжести группы (вектор средних), так как он определяется по оставшимся наблюдениям<sup>5</sup>.

Процедура исключения наблюдений схематично представлена на рис. 2. Она потребовала восьми последовательных переходов (шагов) и продолжалась до тех пор, пока общий коэффициент корректности в классификационной матрице не достиг 100%, значение статистики Уилкса при этом составило 0,0826,

что свидетельствует о хорошей дискриминации итоговой обучающей выборки (рис. 3).

Наличие переменных в моделях оценивалось по уровню толерантности. Значение толерантности определялось как 1 минус *R*-квадрат, где *R*-квадрат - коэффициент множественной корреляции для соответствующей переменной со всеми другими переменными в текущей модели. Он равен доле дисперсии, относящейся к анализируемой переменной.

Согласно итоговой классификационной матрице общее число объектов, оказавшееся в обучающей выборке после исключения всех некорректно отнесенных к классификационным группам объектов, составило 867 единиц (61,45% от первоначального объема совокупности). Распределение этих объектов представлено в табл. 3.

Таблица 3

**Распределение офисных объектов итоговой обучающей выборки по ценовым классам**

Ценовой класс по уровню цены, руб. за 1 кв. м	Число объектов	Удельный вес в составе выборки, %
Очень низкая цена	34	4,00
Низкая цена	704	81,02
Высокая цена	125	14,55
Очень высокая цена	4	0,005
<b>Итого</b>	<b>867</b>	<b>100,00</b>

Полученные на последнем этапе (8-й шаг, рис. 2) обучающие выборки позволяют произвести классификацию объектов, не вошедших в итоговую дискриминацию (это 554 объекта, или 38,55%), либо вычислить классификационные значения для вновь экспонируемых на рынке объектов. С этой целью рассчитываются итоговые значения классификационных функций ( $S_i$ ) по формуле

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

где  $i$  - номер совокупности (ценовой группы);  $j = 1, 2, \dots, m$  - номера переменных;  $C_i$  - константа для  $i$ -й совокупности ( $i$ -й классификационной функции);  $W_{ij}$  - веса для  $j$ -й переменной при вычислении показателя классификации для  $i$ -й совокупности (ценовой группы);  $X_j$  - значение факторной переменной  $j$  для соответствующего объекта офисной недвижимости.

Новый объект должен быть отнесен к тому классу (группе), для которого классификационное значение ( $S_i$ ) максимально.

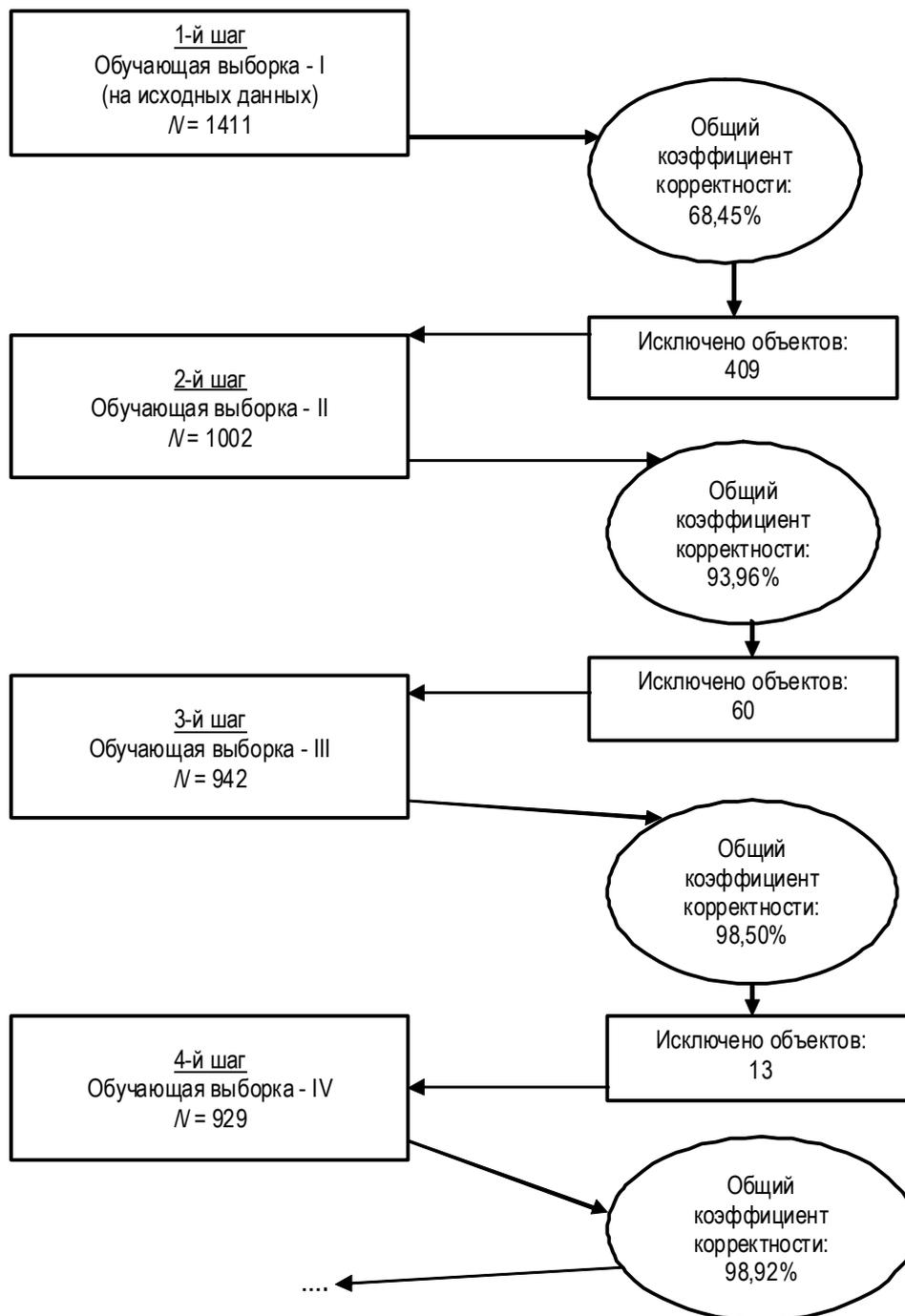


Рис. 2. Алгоритм дискриминантного анализа с последовательным исключением объектов офисной недвижимости из исходной совокупности (начало)

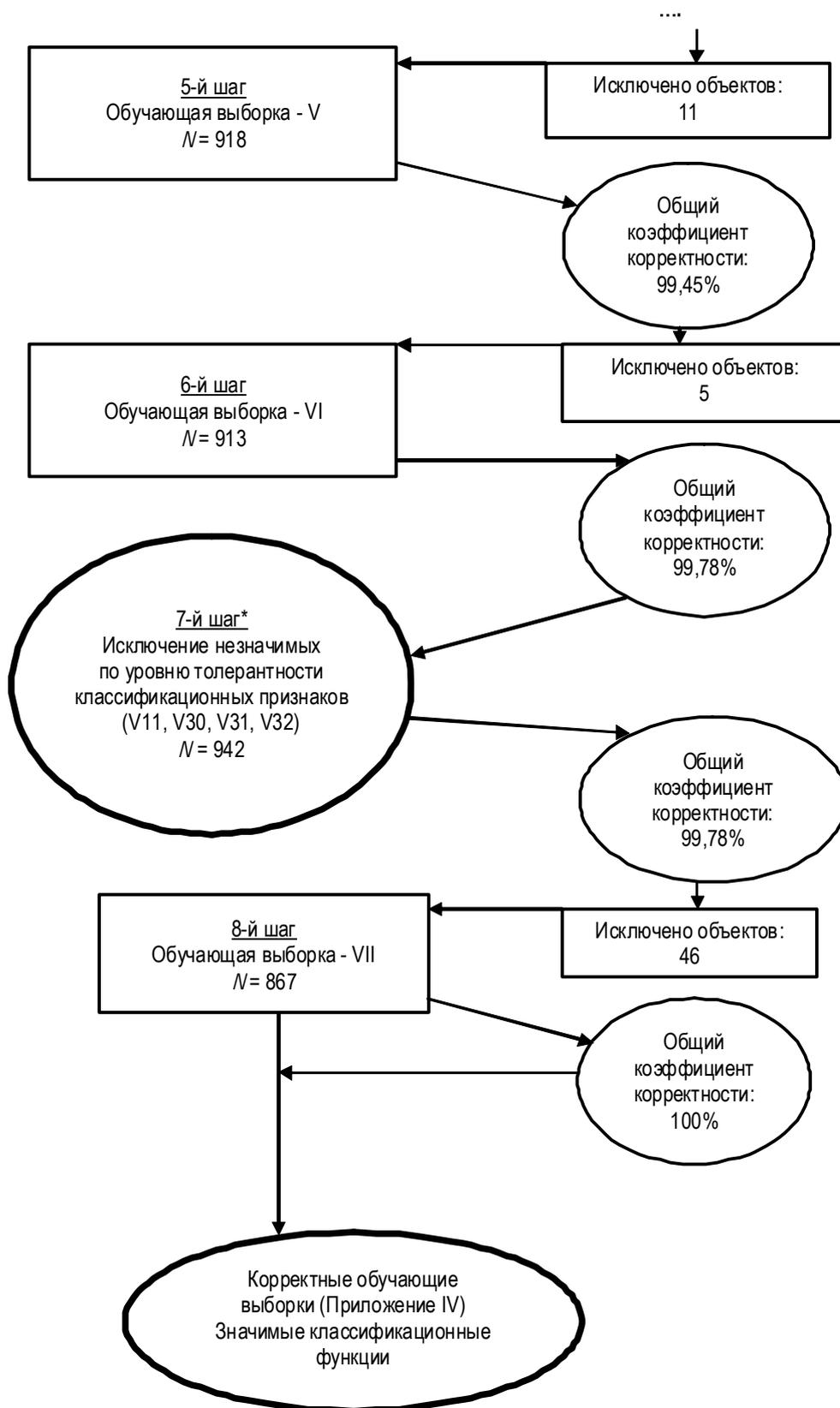
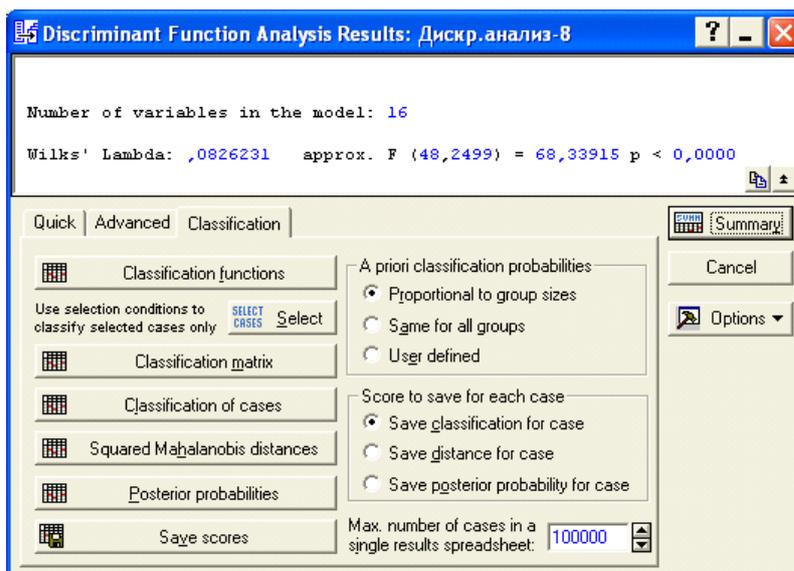


Рис. 2. Алгоритм дискриминантного анализа с последовательным исключением объектов офисной недвижимости из исходной совокупности (окончание)

\* Переменные исключены, так как уровень их толерантности оказался ниже критического (0,01).



**Рис. 3. Результаты анализа дискриминантных функций на завершающем этапе (на 8-м шаге) формирования обучающей выборки распределения объектов офисной недвижимости в г. Самаре по ценовым группам**

Итоговый вид классификационных функций представлен следующими моделями (табл. 4).

На основе полученных обучающих выборок произведена повторная классификация тех объектов офисной недвижимости, которые не попали в обучающие выборки. В результате их перегруппировки в соответствии с рассчитанными классификационными значениями ( $S$ ) полу-

чена итоговая группировка (рис. 4). Единицы, составившие в итоге дискриминантного анализа каждую из групп, обладают общими закономерностями ценообразующего влияния рассматриваемых факторов, комбинация которых и их сила воздействия определяют отнесение объектов к соответствующей ценовой группе. Для исследования этих закономерностей рассчитаны парные коэффициенты корреляции уровня цены

Таблица 4

**Параметры классификационных функций по результатам дискриминантного анализа распределения совокупности объектов офисной недвижимости в г. Самаре на ценовые группы**

	Параметры классификационных функций по ценовым группам			
	Низкая $p=,81024$	Высокая $p=,14552$	Оч.низкая $p=,03958$	Оч.высокая $p=,00466$
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
Константа	-69,2762	-104,973	-74,6900	-140,635

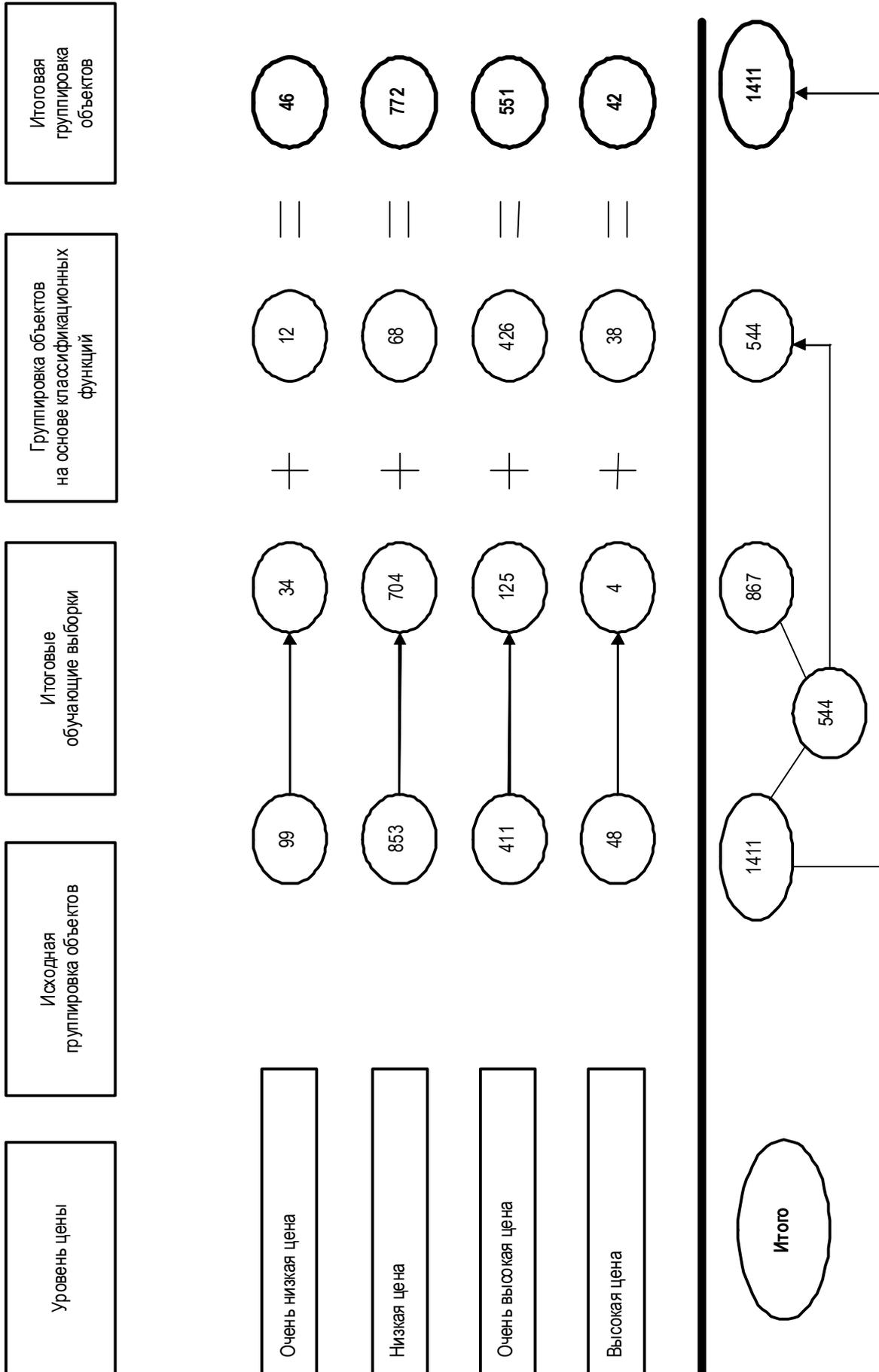


Рис. 4. Схема выполнения дискриминантного анализа распределения объектов офисной недвижимости в г. Самаре по ценовым группам

Таблица 5

**Матрица парных коэффициентов корреляции уровня цены объектов офисной недвижимости в г. Самаре за 1 кв. м (V15) и факторных переменных, обусловивших распределение объектов по ценовым группам**

Наименование переменной, обозначение	Группы объектов по уровню цены			
	Очень низкая	Низкая	Высокая	Очень высокая
Площадь, кв. м - V 1	0,36*	-0,02	-0,17	-0,18
Этаж - V 2	-0,27	-0,12	-0,02	-0,33*
Этажность - V 3	0,51*	0,04	0,05	0,14
Экспозиция, сут - V 4	-0,35*	-0,07	-0,06	-0,14
Наличие рядом остановки (да/ нет) - V 22	0,20	0,10	-0,01	0,05
До ближайшей остановки, м - V 23	0,21	-0,02	0,08	-0,19
До ближайшей улицы, м - V 24	-0,20	-0,07	-0,06	-0,13
До ближайшей главной магистрали, м - V 25	0,33*	-0,03	0,13	-0,13
Интенсивность движения машин, ч - V 26	0,56*	0,10	0,01	-0,10
Маршрутов общего транспорта - V 27	0,85*	0,11	0,04	-0,19
Рейсов общего транспорта в сутки - V 28	0,80*	0,12	0,04	-0,13
Плотность застройки кадастрового квартала - V 29	0,01	0,01	0,04	0,32*
Деловая активность, нет - V 33	0,50*	0,03	0,08	-0,02
Памятников истории - V 34	0,39*	-0,05	0,12	0,08
Является ли памятником истории (да/нет) - V 35	0,02	-0,05	-0,00	0,01
Стоимость жилой недвижимости, руб. за 1 кв. м - V 36	0,40*	0,01	0,25	0,42*

Примечание: \* отмечены парные коэффициенты корреляции, значимые по t-критерию Стьюдента.

объекта офисной недвижимости за 1 кв. м (V15) и факторных переменных, обусловивших дискриминацию объектов по ценовым группам (табл. 5).

Как показывают данные табл. 5, группирующее влияние факторных переменных имеет статистически значимую силу ценообразующего влияния лишь в двух группах: 1 - "Очень низкая цена" (переменные V1, V3, V4, V25, V26, V27, V28, V33, V34, V36) и 2 - "Очень высокая цена" (переменные V2, V29 и V36). По остальным двум группам объектов (с уровнями "Низкая цена" и "Высокая цена") статистически значимого влияния группирующих факторных переменных на уровень цены объекта за 1 кв. м не установлено.

Анализ взаимного влияния факторных переменных с целью устранения мультиколлинеарных связей по данным матриц парных коэффициентов корреляции позволил получить структуру регрессионных моделей цены объектов (V15) по двум группам: "Очень низкая цена" и "Очень высокая цена". В результате оценивания их параметров получены следующие модели:

1. Ценовая группа "Очень низкая цена".

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

где V 15 - последняя цена, руб. за 1 кв. м; V 1 - площадь, кв. м; V 4 - экспозиция, сут; V 26 - интенсивность движения машин, ч; V 27 - маршрутов общего транспорта; V 34 - памятников истории.

Таблица 6

**Параметры и статистические характеристики многофакторной регрессионной модели уровня цены реализации объектов офисной недвижимости, руб. за 1 кв. м (V15) по ценовой группе "Очень низкая цена"**

N=45	Итоги регрессии для зависимой переменной: Var 15 (Дискр.-готово-Оч.низкая) R= ,94644293 R2= ,89575423 Скорректир. R2= ,88238938 F(5,39)=67,023 p					
	БЕТА	Стд.ош.	B	Стд.ош.	t(39)	p-уров.
Св.член			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

Таблица 7

Параметры и статистические характеристики многофакторной регрессионной модели уровня цены реализации объектов офисной недвижимости, руб. за 1 кв. м (V15) по ценовой группе “Очень высокая цена”

N=42	Итоги регрессии для зависимой переменной: Var 15 (Дискр.-готово-Оч.высок) R= ,56426011 R2= ,31838947 Скорректир. R2= ,28343509 F(2,39)=9,1087 p					
	БЕТА	Стд.ош.	В	Стд.ош.	t(39)	p-уров.
Св.член			-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

Статистические характеристики полученного уравнения, подтверждающие его адекватность и значимость параметров, приведены в табл. 6.

Сравнение значений стандартизованных коэффициентов регрессии показателей: отрицательное влияние на уровень цены реализации объектов рассматриваемой группы оказывают факторные переменные: наибольшее по степени - площадь, кв. м (V1) и сравнительно меньшее - время экспозиции на открытом рынке, сут (V4); положительное влияние отказывают: наибольшее - близость памятников истории (V34), среднее по степени - наличие маршрутов общего транспорта (V27), наименьшее - интенсивность движения машин (V26).

Можно сделать обобщение, что уровень цены объектов, относимых к классификационной группе самых дешевых (по цене за 1 кв. м), напрямую зависит от благоустройства и престижности района месторасположения объекта и также находится в существенной прямой зависимости от уровня спроса и, соответственно, в обратной - от конкуренции показателей. Последний для дешевых объектов тем ниже, чем больше их площадь, и его снижение влияет на рост времени экспозиции объектов данной ценовой группы на открытом рынке.

2. Ценовая группа “Очень высокая цена”.

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

где V 15 - последняя цена, руб. за 1 кв. м; V 29 - плотность застройки кадастрового квартала; V 36 - стоимость жилой недвижимости, руб. за 1 кв. м.

Статистические характеристики модели приведены в табл. 7.

Из данных табл. 7 следует, что аппроксимирующие свойства полученной модели невысоки: коэффициент множественной детерминации  $R^2 = 0,318$ , т.е. модель обеспечивает 31,8% объясненной вариации.

Сопоставление значений  $\beta$  (бета)-коэффициентов показывает, что цена 1 кв. м самых дорогостоящих объектов офисной недвижимости в наибольшей степени напрямую зависит от цены 1 кв. м объектов жилой недвижимости (V36) в соответствующем районе и несколько в меньшей степени, но также прямо определяется плотностью застройки кадастрового квартала (V29), в котором этот район расположен.

Таким образом, цена самых дорогостоящих объектов зависит от конкуренции на локальном рынке недвижимости микроуровня в городском масштабе и общего ценового уровня этого рынка.

Незначимость корреляционных связей по двум другим ценовым группам исследуемой совокупности офисных объектов (“Низкая цена” и “Высокая цена”), а также необходимость углубления исследования ценообразующего влияния факторов приводит к необходимости перехода к следующему этапу дискриминантного анализа - каноническому анализу.

В комплексе предложенная методика позволяет выработать адекватную статистическую панель для построения многофакторных регрессионных моделей ценообразования на объекты офисной недвижимости города, исходя из содержательной однородности условий и факторов обеспечения формирования цены.

<sup>1</sup> Дуброва Т.А., Бажин А.Г., Бакуменко Л.П. Дискриминантный анализ в системе “STATISTICA”: Учеб.пособие. М., 2000. С. 4.

<sup>2</sup> Там же. С. 26.

<sup>3</sup> Там же. С. 28.

<sup>4</sup> Халафян А.А. STATISTICA 6. Статистический анализ данных: Учебник. 3-е изд. М., 2008. С. 203-204.

<sup>5</sup> Дуброва Т.А., Бажин А.Г., Бакуменко Л.П. Указ. соч. С. 51.