

Крапивина Анастасия Алексеевна,

студент кафедры прикладной математики, специальность «Математическое и информационное обеспечение в экономической деятельности»

ФГБОУ ВПО «Пермский национальный исследовательский политехнический университет», г. Пермь

ФАКТОРНЫЙ И КЛАСТЕРНЫЙ АНАЛИЗ РЫНКА ВТОРИЧНОГО ЖИЛЬЯ

Аннотация. В статье анализируется рынок вторичного жилья Пермского края. При помощи факторного и кластерного анализа определена степень влияния факторов, характерных для вторичного жилья, на цену однокомнатной квартиры.

Ключевые слова: факторный анализ, кластерный анализ, цена, вторичное жилье.

Abstract. The article analyzes the secondary housing market in Perm region. With the help of factor and cluster analysis the degree of influence of factors, characteristic for the secondary housing, the price of a studio apartment.

Keywords: factor analysis, cluster analysis, the price of secondary housing.

При продаже (покупке) квартиры важно иметь представление о ее оптимальной стоимости в условиях текущего рынка. Для оценки стоимости квартиры можно использовать статистический анализ рынка.

К рассмотрению представлен вторичный рынок однокомнатных квартир города Перми [1].

Исходные данные представлены в виде списка из 70 однокомнатных квартир со следующими характеристиками: цена квартиры, площадь общая, площадь жилая, площадь кухни и следующими экспертными оценками: удаленность от центра, этаж квартиры, тип дома, состояние квартиры. За результирующий показатель выборки принята цена квартиры.

Определена следующая градация по переменным:

- Y – цена квартиры (тыс.руб.);
- X_1 – общая площадь квартиры (m^2);
- X_2 – жилая площадь квартиры (m^2);

- X3 – площадь кухни (м²);
- X4 – тип дома:
 1. малосемейка;
 2. хрущевка;
 3. брежневка;
 4. полногабаритная (сталинка);
 5. улучшенной планировки;
- X5 – этаж:
 1. крайние этажи;
 2. средние этажи;
- X6 – удаленность от центра:
 1. центр (Драмтеатр, Компрос);
 2. приближенные к центру (Парковый, Садовый, Балатово, Пролетарский, Горки, комплекс);
 3. средняя удаленность (Крохалева, Заостровка, Железнодорожный, Рабочий поселок, Нагорный);
 4. удаленные (Вышка-2);
 5. самые удаленные (Гайва, Левшино, Закамск, Крым, Голованово);
- X7 – состояние квартиры:
 1. без ремонта;
 2. обычное состояние;
 3. евроремонт.

Проведем факторный анализ введенных переменных. Матрица собственных значений [2] приведена на рис. 1.

	Соб. зн.	% общей	Кумулятивн.	Кумулятивн.
1	3,534393	50,49132	3,534393	50,4913
2	1,224797	17,49711	4,759190	67,9884
3	0,997835	14,25479	5,757025	82,2432
4	0,562264	8,03235	6,319290	90,2756
5	0,500997	7,15711	6,820287	97,4327
6	0,123742	1,76775	6,944029	99,2004
7	0,055971	0,79958	7,000000	100,0000

Рис. 1 – Матрица собственных значений. Выделение главных компонент

Исходя из полученной матрицы собственных значений матрицы парных корреляций, наибольший вклад в суммарную дисперсию системы внесли компоненты F1 и F2, так как значение их собственных чисел больше единицы и суммарный вклад около 68%. Причем первая главная компонента объясняет 50,5%, вторая – 17,5% суммарной дисперсии системы соответственно.

Из рис. 2 очевидно, что фактор 1 (привлекательность квартиры) явно зависит от общей площади квартиры (X1), жилой площади (X2), площади кухни (X3), типа дома (X4). Включим также в фактор 1 показатель X7 – состояние квартиры. Фактор 2 зависит от удаленности квартиры от центра (X6).

	Фактор	Фактор	Фактор	Фактор	Фактор	Фактор	Фактор
X1 общая площадь	0,944301	0,092587	0,024276	-0,101118	-0,092370	0,273350	0,075210
X2 жилая площадь	0,808529	0,004641	0,066771	-0,197587	-0,530137	-0,145731	-0,021840
X3 площадь кухни	0,939387	0,090719	-0,000693	0,214912	0,170208	0,026282	-0,182958
X4 тип дома	0,879737	0,133448	0,053872	0,366198	0,185456	-0,143706	0,127301
X5 Этаж	0,092143	0,037617	-0,993326	0,031676	-0,047935	-0,007062	0,006818
X6 удаленность от центра	0,046754	0,911007	-0,006491	-0,360438	0,188320	-0,049564	-0,001349
X7 состояние	0,567368	-0,599013	-0,056084	-0,449213	0,332104	-0,062736	0,010485
Общая дисперсия	3,534393	1,224797	0,997835	0,562264	0,500997	0,123742	0,055971
Доля общей дисперсии	0,504913	0,174971	0,142548	0,080323	0,071571	0,017677	0,007996

Рис. 2 – Матрица факторных нагрузок (без вращения). Выделение главных компонент

Построим уравнение регрессии производительности труда (Y) на главной компоненте [3]. Методом пошаговой регрессии получим следующие результаты (рис. 3).

	БЕТА	Стд.Ош.	В	Стд.Ош.	t(67)	p-уров.
Свободный член			2066.271	18.46508	111.9015	0.000000
Фактор 1	0.853655	0.039573	401.202	18.59841	21.5718	0.000000
Фактор 2	-0.407861	0.039573	-191.687	18.59841	-10.3066	0.000000

Рис. 3 – Итоги регрессии для зависимой переменной: Y – цена

Уравнение регрессии в размерной форме имеет вид:

$$Y = 2066,27 + 401,20 * \text{Фактор 1} - 191,69 * \text{Фактор 2}.$$

Фактор 1 в большей степени влияет на Y. Коэффициент множественной детерминации высокий ($R_0^2 = 0,895$), следовательно, уравнение значимо.

Коэффициенты уравнения регрессии показывают, на сколько единиц может измениться результат при изменении факторов на 1 единицу.

При изменении фактора 1 на 1 единицу цена квартиры может измениться на 401,20 единиц, при изменении фактора 2 на единицу – может уменьшиться на 191,69 единиц.

Проиллюстрируем применение полученной регрессии на примере квартиры брежневского типа (X4) в центре города (X6) с общей площадью (X1) 30 м², жилой площадью (X2) 18 м², площадью кухни (X3) 6 м², на крайнем этаже (X5), в обычном состоянии (X7).

Уравнение регрессии в размерном виде примет следующий вид:

$$\begin{aligned} Y &= 2066,27 + 401,20 * \text{Фактор 1} - 191,69 * \text{Фактор 2} = \\ &= 2066,27 + 401,2 * (-0,567) - 191,69 * (-1,334) = 2094,5 \end{aligned}$$

Реальная цена квартиры: 1990 тыс.руб.

Цена, полученная при расчете, близка к реальной, но есть погрешность, потому как с помощью уравнения регрессии можно описать 94,6% дисперсии ($R_0^2 = 0,946$).

Построим доверительный интервал прогноза:

$$1989,8 \leq Y \leq 2199,2.$$

Реальная цена квартиры: 1990 тыс. руб., она попадает в доверительный интервал прогноза. Если цена квартиры находится ближе к левому концу интервала, то она считается заниженной. Если же цена ближе к правому концу – считается завышенной. Заниженная цена вызывает опасения у покупателей, а квартиру с неоправданно завышенной ценой довольно сложно продать. Поэтому обычно указывается цена чуть выше средней с учетом торга.

Объекты расположились в пространстве двух главных факторов следующим образом (рис. 4):

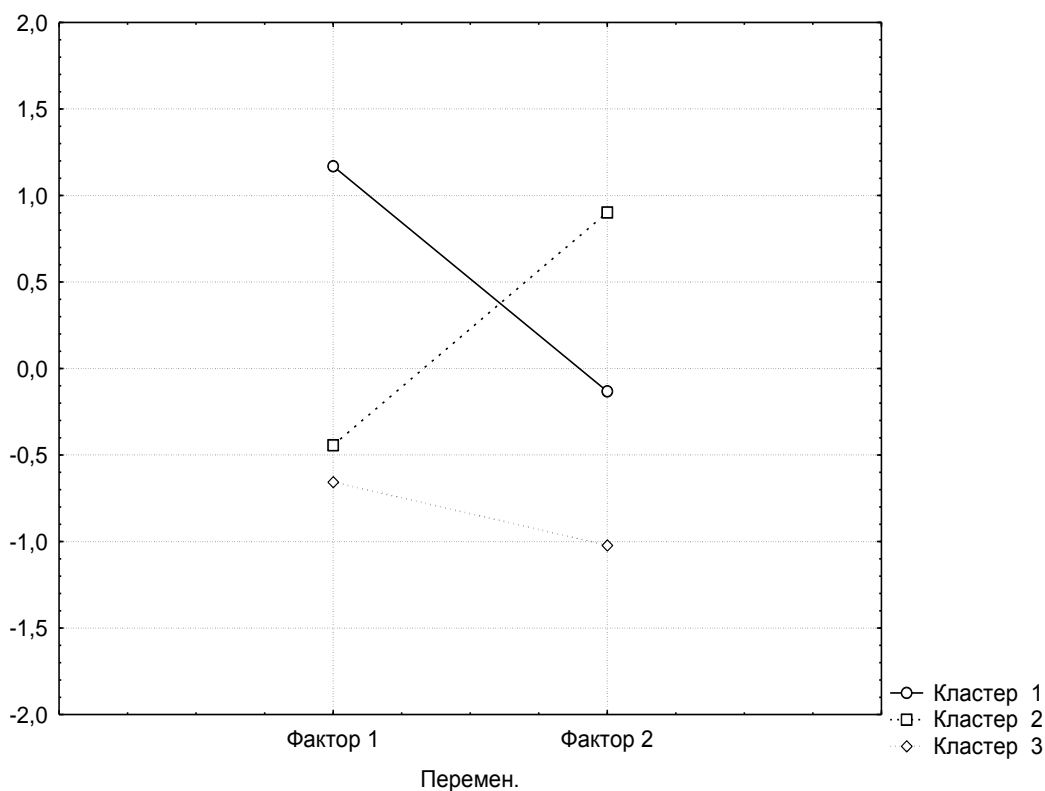


Рис. 4 – График средних для каждого кластера

Получившиеся кластеры можно охарактеризовать следующим образом:

- Кластер 1 (Высший класс): привлекательные для покупки квартиры, важна не близость к центру, а отличное состояние квартиры.
- Кластер 2 (Средний класс): квартиры расположены ближе к центру, состояние обычное, выше среднего.
- Кластер 3 (Низший класс): состояние квартир среднее или ниже, важна близость к центру.

Проведем кластерный анализ введенных переменных.

Произвели нормировку числовых данных. Сгруппировали данные в 3 кластера.

Для первой нормировки будем использовать метод k-средних [4].

Построим график средних (рис. 5) и проанализируем дисперсии (рис. 6):

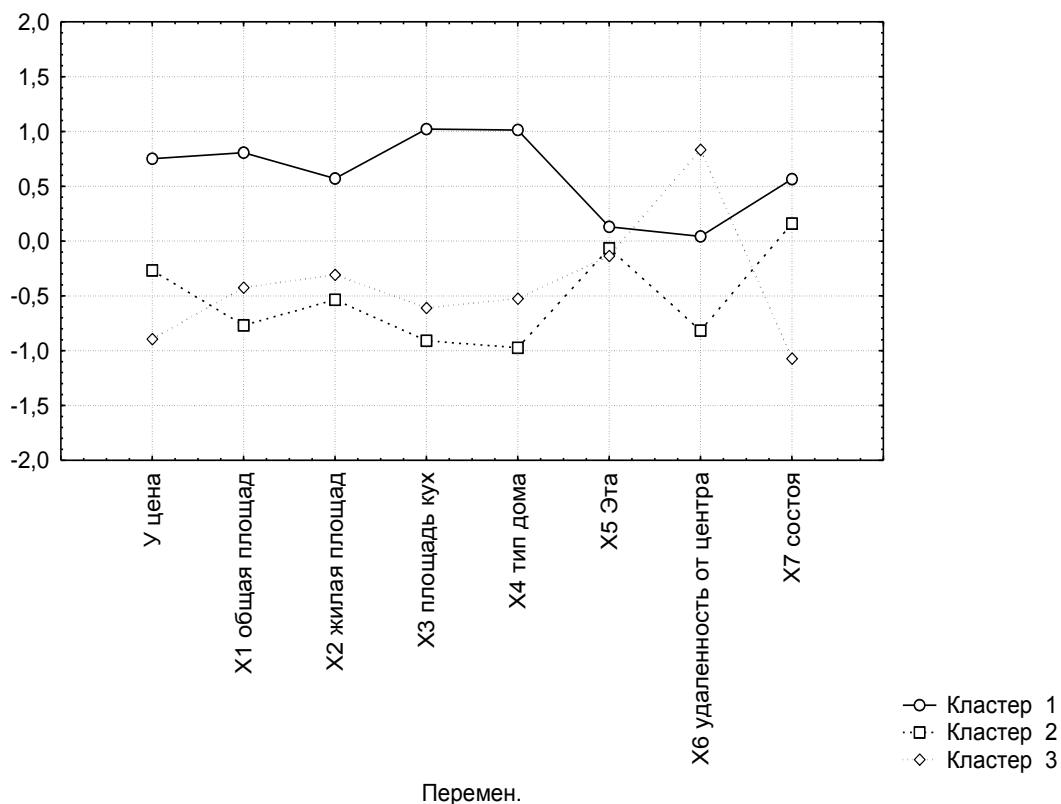


Рис. 5 – График средних для каждого кластера

	Между	сс	Внутри	сс	F	значим.
Y цена	33,54130	2	35,45870	67	31,6885	0,000000
X1 общая площадь	35,38176	2	33,61824	67	35,2573	0,000000
X2 жилая площадь	17,58580	2	51,41420	67	11,4584	0,000052
X3 площадь кухни	55,70234	2	13,29766	67	140,3276	0,000000
X4 тип дома	55,84536	2	13,15464	67	142,2175	0,000000
X5 Этаж	0,93117	2	68,06882	67	0,4583	0,634340
X6 удаленность от центра	27,29764	2	41,70236	67	21,9285	0,000000
X7 состояние	31,92043	2	37,07957	67	28,8389	0,000000

Рис. 6 – Дисперсионный анализ

По трем признакам межклассовая дисперсия больше внутриклассовой дисперсии, т.е. значения этих признаков в кластерах близки между собой.

Квартиры были распределены по кластерам следующим образом: в кластер 1 попало двадцать две квартиры, в кластер 2 – двадцать семь, а в кластер 3 – двадцать одна квартира.

В результате проведения факторного анализа объединены показатели в два фактора: Фактор 1 (привлекательность квартиры для покупки) включает в себя цену.

площадь общую, площадь жилую, площадь кухни, тип дома и состояние квартиры. Фактор 2 (удаленность от центра) включает в себя удаленность от центра. Построено также уравнение зависимости между ценой квартиры и двумя факторами.

В результате проведения кластерного анализа разбита выборка объектов на 3 группы (кластера). Отличительной чертой первого кластера является высокая цена квартиры при отличном состоянии с возможной удаленностью от центра, отличительной чертой второго кластера является средняя цена квартиры при хорошем состоянии (квартиры приближены к центру), отличительной чертой третьего кластера является низкая цена квартиры при среднем или ниже среднего состоянии с максимальной близостью к центру.

Библиографический список

1. Продажа квартиры на вторичном рынке [Электронный ресурс]. Режим доступа: www.dom.59.ru (Дата обращения: 10.04.2016).
2. Факторный, дискриминантный и кластерный анализ / сборник работ под ред. Енюкова И. С. — М.: Финансы и статистика, 1989. — 215 с.
3. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Множественная регрессия = Applied Regression Analysis. — 3-е изд. — М.: «Диалектика», 2007. — 912 с.
4. Мандель И. Д. Кластерный анализ. — М.: Финансы и статистика, 1988. — 176 с.