

УДК 332.64

DOI:10.24412/2782-4845-2023-5-65-74

ПРИМЕНЕНИЕ КОРРЕЛЯЦИОННО–РЕГРЕССИОННОГО ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА СОЦИАЛЬНО – ЗНАЧИМЫХ ФАКТОРОВ

Е.А. Рязанцева, Липецкий филиал ФГОБУ ВО «Финансовый университет при Правительстве РФ», Липецк, Россия

***Аннотация.** Статья посвящена исследованию взаимосвязи показателя «стоимость квартиры» с различными факторами, влияющими на его формирование. Используя возможности табличного редактора, позволяющего быстро получать решение, был проведен корреляционно-регрессионный анализ. Данные были взяты из открытых источников по Липецкой области за период 2014–2021 годы, показатели рассматриваются в среднем. В результате проведенного исследования был сделан вывод, что наиболее значимым фактором, который влияет на стоимость квартиры, остается площадь квартиры. Полученная множественная регрессия отвечает требованиям, которые предъявляются экономико-математическим моделям и, с учетом значения ошибки аппроксимации, может быть использована для дальнейшего исследования и прогнозирования.*

***Ключевые слова:** множественная регрессия, корреляционно-регрессионный анализ, стоимость квартиры в Липецкой области.*

Для цитирования: Рязанцева Е.А. Применение корреляционно-регрессионного инструментария для анализа социально-значимых факторов // ЭФО: Экономика. Финансы. Общество. 2023. № 1 (5). С.65-74. DOI:10.24412/2782-4845-2023-5-65-74

THE USE OF CORRELATION-REGRESSION TOOLS FOR THE ANALYSIS OF SOCIALLY SIGNIFICANT FACTORS

E.A. Ryazantseva, Lipetsk branch of the Financial University under the Government of the Russian Federation, Lipetsk, Russia

***Abstract.** The paper is devoted to the study of the interrelation between the indicator "cost of an apartment" and various factors, which can affect the cost. Using the capabilities of the tabular editor, which allows you to get a solution much faster, a correlation-regression analysis was carried out. The data were taken from open sources in the Lipetsk region and covered the period 2014-2021, indicators are considered on average. The study showed that the most significant factor, which affects the cost of the apartment remains the area of the apartment. The resulting multiple regression meets the requirements of economic and mathematical models and, taking into account the average relative error, can be used for the further research and forecasting.*

***Keywords:** multiple regression, correlation-regression analysis, cost of an apartment in the Lipetsk region.*

Введение

Исторически сложилось, что деятельность рынка недвижимости подвержена воздействию ряда факторов, которые предопределяют возникновение крупных кризисов. Можно предположить о недостаточном внимании аналитиков при изучении данной предметной области. В процессе ценообразования жилой недвижимости просматривается влияние многих факторов, как явных, так и скрытых.

Первый и важный фактор, который в большинстве случаев влияет на конечную стоимость жилья – это географическое положение. Как показывает анализ предметной области на основании последних исследований [2] – [4], самая высокая стоимость наблюдается в крупных мегаполисах, таких как Москва, Санкт-Петербург, Екатеринбург, Сочи и других. Стоимость аналогичных квартир в региональных городах существенно ниже. Объяснение данного факта простое – это существенно различный уровень жизни, обусловленный, в том числе, и уровнем зарплат населения [2].

Также можно выделить год постройки, материал, из которого построен данный дом, и этажность. Каждый из данных факторов в той или иной степени влияет на конечную стоимость жилой недвижимости [4].

В данном исследовании был выбран регион – Липецкая область, который расположен в центральной части Российской Федерации на пересечении важнейших транспортных артерий, к примеру М4. Численность населения в области составляет $\frac{5}{4}$ миллиона человек. Более 60% валового регионального продукта относится к промышленному производству. Кроме того, создана экономическая особая зона, которая активно развивается. На территории области расположены около двухсот предприятий, которые осуществляют свою деятельность в различных сферах деятельности. Можно выделить ряд крупнейших, такие как ПАО «Новолипецкий металлургический комбинат» ЗАО Завод «Росинка», ОАО «Лебедянский» и многих других.

Важно отметить, что Липецкая область, хоть и молодая (дата образования – 6 января 1954 года), но имеет огромный потенциал и ресурсы для развития. Здесь есть и рабочие места, и возможности профессионального роста, возможности приобретения жилья, как на вторичном рынке, так и на рынке новостроек [9].

Особое внимание привлекает Липецк как областной центр. В нем в последние годы наблюдается активное строительство новых микрорайонов – 20-28 микрорайоны. Все квартиры данных районов улучшенной планировки и возраст домов менее двадцати лет. Также в историческом центре города возводятся жилищные комплексы, которые имеют повышенный спрос.

Изучение эмпирических данных удобнее всего вести, если есть возможность представить данные в аналитическом виде. Прежде всего, функциональная зависимость позволяет проводить необходимый анализ, строить прогноз, а также успешно принимать решения.

Таким образом, была поставлена задача об изучении зависимости

стоимости квартиры в городе Липецке. При исследовании были выбраны такие факторы, которые могли бы влиять на данный результирующий признак: район города, год постройки, материал, площадь, этажность.

Основная часть

Необходимые данные взяты из открытых источников <https://lipstat.gks.ru/> – Территориальный орган Федеральной службы государственной статистики по Липецкой области. В таблице 1 представлены исходные данные для анализа.

Таблица 1. Статистические данные исследуемых показателей*

Стоимость квартиры, руб.	Район города	Год постройки	Материал	Площадь, м ²	Этажность
y	x_1	x_2	x_3	x_4	x_5
1 950 000	4	1960	1	36	4
4 500 000	3	1997	1	55	6
3 970 000	1	2010	3	62,1	21
7 000 000	1	2019	1	61	22
6 950 000	1	2000	1	126,3	6
5 200 000	2	2006	4	80	9
4 230 000	2	2016	4	51	10
4 500 000	4	1960	7	73,3	4
2 470 000	1	1966	5	32,3	5
3 980 000	1	2016	2	40,5	21
4 450 000	1	1999	2	53,2	10
3 380 000	1	2018	2	34,5	25
3 300 000	1	2015	2	42,3	19
7 290 000	1	2021	2	76	13
7 700 000	1	2019	3	56	25
7 950 000	1	2015	1	77,6	19
6 700 000	2	1979	1	54	5
6 600 000	3	2016	2	70,8	16
5 250 000	3	2002	4	67,9	10
3 000 000	4	1958	1	55,9	5
1 580 000	4	1975	1	25,1	5

*составлено автором по данным [9]

В данном случае, для построения модели [10] необходимо, чтобы в качестве объясняющих переменных выступали как численные величины (стоимость, год постройки, этажность, площадь), так и качественные переменные (материал, район города). Обычно в роли таких переменных выступают дихотомические или, по-другому, бинарные переменные или фиктивные. Они позволяют отслеживать структурные изменения в анализируемых явлениях или процессах. При этом необходимо иметь в виду, что количество таких переменных должно быть на единицу меньше, чем число уровней изучаемого признака.

Таким образом, в модель будут введены следующие количественные

переменные – стоимость квартиры (руб.), год постройки, этажность, площадь (м²). Остальные же параметры (материал, район города) отражают качественные характеристики, поэтому они будут являться фиктивными и принимать значения, указанные в таблицах 2 и 3.

Таблица 2. Числовые обозначения переменной «Материал»*

Значения	Вид фактора «Материал»
1	Кирпичный
2	Монолитный
3	Кирпично-монолитный
4	Панельный
5	Блочный
6	Деревянный
7	Сталинский

**составлено автором на основе расчетов*

Таблица 3. Числовые обозначения переменной «Район города»*

Значения	Вид фактора «Район города»
1	Советский
2	Правобережный
3	Октябрьский
4	Левобережный

**составлено автором на основе расчетов*

Важнейшим элементом эконометрического исследования является графический анализ исходных данных. Для создания диаграммы рассеяния нужно выделить два столбца данных со значениями показателей, включая их названия (метки) в первой строке матрицы данных (рис.1).

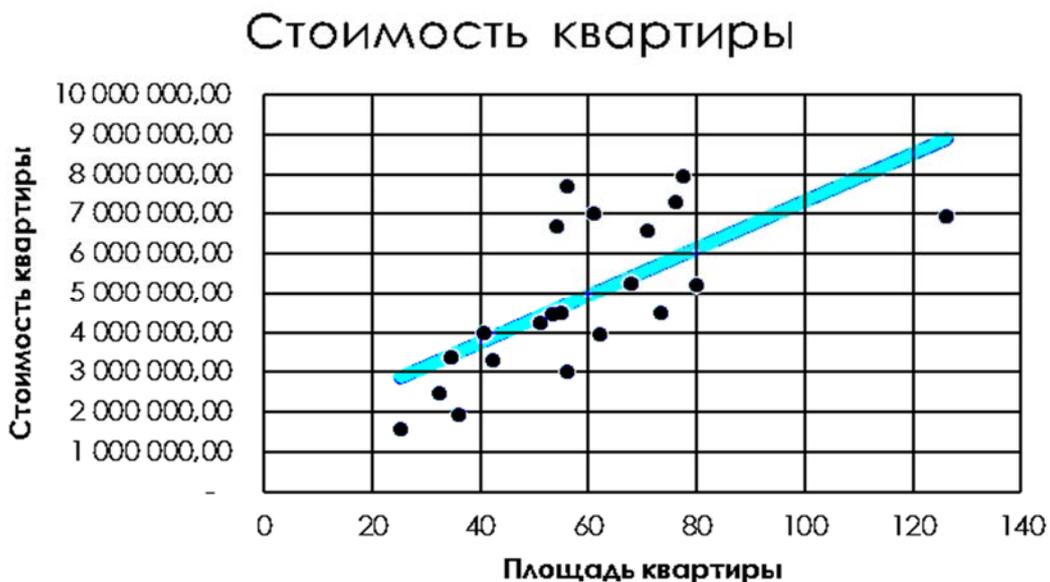


Рис.1. Зависимость переменной стоимость квартиры от фактора «площадь квартиры»*

**составлено автором на основе расчетов*

Исходя из представленных данных, диаграмма рассеяния имеет вид, приведенный на рис. 1. Вытянутость облака точек на диаграмме рассеяния вдоль наклонной прямой позволяет сделать предположение о том, что существует некоторая объективная тенденция прямой линейной связи между значениями переменных x и y .

Для проведения корреляционного анализа используем надстройкой «Корреляция». В результате будет получена матрица коэффициентов парной корреляции (см. рис. 2).

Матрица парных коэффициентов корреляции						
	Y	X1	X2	X3	X4	X5
Y	1					
X1	-0,423514963	1				
X2	0,549724001	-0,688678636	1			
X3	-0,119809373	0,075269301	-0,158710793	1		
X4	0,683313642	-0,161305699	0,223119941	0,027090942	1	
X5	0,341855892	-0,641185316	0,795842994	-0,131073595	-0,083535692	1

Рис. 2. Результат воздействия надстройки «Корреляция»*

*составлено автором на основе расчетов

Исследование матрицы коэффициентов парной корреляции начнем с анализа первого столбца матрицы, в котором расположены коэффициенты корреляции, отражающие тесноту связи зависимой переменной «Стоимость квартиры» с включенными в анализ факторами. Анализ показывает, что зависимая переменная, то есть стоимость квартиры, имеет тесную связь с годом постройки (0,54) и площадью (0,68). Факторы x_1 , x_2 , x_3 имеют слабую связь с зависимой переменной, поэтому их не рекомендуется включать в модель регрессии.

Оценим значимость коэффициентов корреляции первого столбца матрицы. Для этого рассчитаем значение t – статистики для всех элементов первого столбца по формуле (1):

$$t_{ij} = \frac{r_{ij0}\sqrt{n-k-1}}{\sqrt{1-r_{ij0}^2}} \quad (1)$$

Используя возможности MS Excel были осуществлены расчеты (2):

$$\begin{aligned} t_{yx_1} &= -2,09; \\ t_{yx_2} &= 2,94; \\ t_{yx_3} &= -0,54; \\ t_{yx_4} &= 4,19; \\ t_{yx_5} &= 1,63. \end{aligned} \quad (2)$$

Табличное значение критерия Стьюдента равно: $t_{\text{табл}}(0,05, 20) = 2,086$. Сравним полученные значения t -критериев с табличным. Таким образом, установлено, что $t_{yx_2} > t_{\text{табл}}$ и $t_{yx_4} > t_{\text{табл}}$, т.е. полученные значения коэффициентов корреляции значимы для года постройки и площади.

Значимость коэффициентов корреляции можно проверить, используя критическое значение коэффициента корреляции. При условии, что нулевая гипотеза $H_0: r_{ij} = 0$, критическое значение коэффициента корреляции определяется статистикой по формуле (3):

$$r^* = \sqrt{\frac{t_{(\alpha, n-2)}^2 / (n-2)}{1 + (t_{(\alpha, n-2)}^2 / (n-2))}} \quad (3)$$

где $t_{(\alpha, n-2)}^2$ – критическое значение t-статистики Стьюдента для уровня значимости α и количества степеней свободы, равного $n-2$.

Так как в рассматриваемом исследовании критическое (табличное) значение критерия Стьюдента равно 2,086, то критическое значение коэффициента корреляции будет равно 0,423. Все коэффициенты парной корреляции в анализируемой матрице, превышающие значение 0,423 по абсолютной величине будут значимы.

Затем перейдем к анализу остальных столбцов матрицы с целью выявления коллинеарности. Одним из условий регрессионной модели является предположение о линейной независимости объясняющих переменных, т. е. решение задачи возможно лишь тогда, когда столбцы и строки матрицы исходных данных линейно независимы. Для экономических показателей это условие выполняется не всегда. Под мультиколлинеарностью понимается высокая взаимная коррелированность объясняющих переменных, которая приводит к линейной зависимости нормальных уравнений. Один из подходов определения наличия или отсутствия мультиколлинеарности заключается в анализе матрицы коэффициентов парной корреляции. Считают явление мультиколлинеарности в исходных данных установленным, если коэффициент парной корреляции между двумя переменными больше 0,8. Анализ показывает, что зависимая переменная, то есть стоимость квартиры, имеет тесную связь с годом постройки (0,54) и площадью (0,68).

Таким образом, на основе анализа корреляционной матрицы для включения в модель регрессии остаются два фактора – «Год постройки» и «Площадь» ($n = 22$, $k = 2$), т.е. в виде функциональной зависимости (4):

$$y = f(x_2, x_4) \quad (4)$$

Для проведения регрессионного анализа используем инструмент Регрессия (надстройка Анализ данных в Excel). Результат вычислений представлен на рисунке 3.

Используя протокол регрессионного анализа, уравнение зависимости стоимости квартиры от года постройки и площади можно записать в следующем виде:

$$y_i = -70342707,6 + 36119,6x_2 + 51443,4x_4$$

Вывод итогов				
Регрессионная статистика				
Множественный R		0,79561529		
R-квадрат		0,63300369		
Нормированный R-квадрат		0,592226322		
Стандартная ошибка		1232951,97		
Наблюдения		21		
Дисперсионный анализ				
	df	SS	MS	F
Регрессия	2	4,71965E+13	2,35982E+13	15,52340735
Остаток	18	2,73631E+13	1,52017E+12	
Итого	20	7,45595E+13		
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
У-пересечение	-70342707,6	25135624,9	-2,8	0,0
X2	36119,6	12655,2	2,9	0,0
X4	51443,4	12771,4	4,0	0,0

Рис. 3. Результат выполнения регрессии*

**составлено автором на основе расчетов*

Получено уравнение регрессии, все коэффициенты которого значимы при 5%-ном уровне значимости. В задаче величина, равная 36119,6 (коэффициент при x_2), показывает, что при увеличении года постройки на 10 лет стоимость квартиры увеличится на 361 196 руб., а если на десять квадратных метров увеличится площадь, то стоимость квартиры увеличится на 514 434 руб.

Для оценки качества модели множественной регрессии вычисляют коэффициент детерминации R^2 и коэффициент множественной корреляции (индекс корреляции) R , чем ближе к 1 значение этих характеристик, тем выше качество модели. А также точность модели оценивается с помощью средней ошибки аппроксимации. Результаты вычислений:

1. Коэффициент детерминации $R^2 = 0,633$ – показывает, что доля вариации результативного признака под воздействием изучаемых факторов составляет 63,3%.

2. Коэффициент множественной корреляции $R = 0,796$ – показывает высокую тесноту связи зависимой переменной с двумя включенными в модель факторами.

3. Средняя ошибка аппроксимации $E_{отн} = 18,77\%$, что говорит о неточности модели.

Проверку значимости уравнения регрессии произведем на основе F-критерия Фишера по формуле:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}, \quad (5)$$

где, k – количество факторов, включенных в модель;

n – количество измерений.

Результат вычисления представлен на рисунке 4.

ОЦЕНКА ЗНАЧИМОСТИ УРАВНЕНИЯ РЕГРЕССИИ И ЕГО КОЭФФИЦИЕНТОВ			
Проверка значимости уравнения регрессии	Ф-критерий Фишера ($F_{\text{расчет}}$)	16,4	Поскольку $F_{\text{расч}} > F_{\text{табл}}$, уравнение регрессии следует признать значимым, то есть его можно использовать для анализа и прогнозирования.
	$F_{\text{табл}}$	3,52	
	$\alpha =$	0,05	
	$\nu_1 = k =$	2	
	$\nu_2 = n - k - 1 =$	19	

Рис. 4. Результат вычислений оценки значимости уравнения регрессии и его коэффициентов*

**составлено автором на основе расчетов*

Табличное значение F-критерия при доверительной вероятности $\alpha = 0,05$ и числе степеней свободы, равном $\nu_1 = k = 2$ и $\nu_2 = n - k - 1 = 22 - 2 - 1 = 19$ составляет 3,52.

Поскольку $F_{\text{расч}} > F_{\text{табл}}$, уравнение регрессии следует признать значимым, то есть его можно использовать для анализа и прогнозирования.

Оценим силу связи факторов с результатом с помощью коэффициентов эластичности, β - и Δ -коэффициентов.

Учитывая, что коэффициент регрессии невозможно использовать для непосредственной оценки влияния факторов на зависимую переменную из-за различия единиц измерения и разной колеблемости факторов, используем коэффициенты эластичности и бета-коэффициенты (рис. 5).

Коэффициент эластичности показывает, на сколько процентов изменяется зависимая переменная при изменении фактора на 1 %.

Показатель	Формула		Расчет
Коэффициент эластичности	$\mathcal{E}_j = a_j \times \bar{x}_j / \bar{y}$	\mathcal{E}_2	14,8684
		\mathcal{E}_4	0,6211
Бета-коэффициент	$\beta_j = a_j \times S_{x_j} / S_y$	β_2	0,418
		β_4	0,590
Дельта-коэффициент	$\Delta_j = r_{y,x_j} \cdot \beta_j / R^2$	Δ_2	0,363
		Δ_4	0,637

Рис. 5. Результат расчета оценки влияния факторов на объем реализации*

**составлено автором на основе расчетов*

Бета-коэффициент с математической точки зрения показывает, на какую часть величины среднеквадратического отклонения меняется среднее значение

зависимой переменной с изменением независимой переменной на одно среднеквадратическое отклонение при фиксированных на постоянном уровне значениях остальных независимых переменных. При увеличении года постройки на 22,35 стоимость квартиры увеличится на 41,8 тыс. руб.

Заключение

Регрессионный анализ остается одной из наиболее востребованных и популярных количественных методов в социальных науках. Возможность одновременного изучения неограниченного количества объектов, а также «прозрачность» техники создали репутацию надежного инструмента анализа. Сильная сторона метода состоит в том, что он направлен не просто на изучение изменений, а на сведение причины и следствия. Иначе говоря, регрессионный анализ отвечает на вопрос: «Влияет ли одна или несколько переменных (потенциальных причин) на другую переменную (результат) и, если да, то в какой степени?». Таким образом, исходя из проведенных расчетов и полученных результатов, можно сделать вывод, что на стоимость квартиры большее влияние оказывает фактор «площадь».

Список использованных источников:

1. Батраченко Е.К. Прогнозирование стоимости квадратного метра в новостройках в городе Самаре // Теория и практика общественного развития. 2019. № 7 (137). С. 33-36.
2. Бойко Т.А. Разработка алгоритма построения модели множественной линейной регрессии // Экономика и управление: проблемы, решения. 2018. Т. 8. № 12. С. 113-119.
3. Бурова И. В. Использование регрессионного анализа в оценке стоимости объектов регионального рынка недвижимости // РППЭ. 2020. №2 (112). URL: <https://cyberleninka.ru/article/n/ispolzovanie-regressionnogo-analiza-v-otsenke-stoimosti-obektov-regionalnogo-rynka-nedvizhimosti> (дата обращения: 12.12.2022).
4. Вертегел С.В. Анализ взаимосвязи цены недвижимости г. Москвы и влияющих на нее факторов // Хроноэкономика. 2020. № 5 (26). С. 47-49.
5. Мамченко О.П., Исаева О.В. и др. Эконометрические методы исследования рынка недвижимости // Управление экономическими системами: электронный научный журнал. 2019. № 5 (123). С. 10.
6. Митяков С.Н., Митяков Е.С., Митякова О.И., Яковлева Г.Н. Инструментарий оценки инновационной деятельности в регионах: корреляционно-регрессионный анализ // Инновации. 2021. № 1 (267). С. 60-67.
7. Пожидаева Д.Ю. О применении модели множественной регрессии в исследовании среднедушевых доходов населения Российской Федерации // Современные подходы к трансформации концепций государственного регулирования и управления в социально-экономических системах. Сборник

научных трудов 10-й Международной научно-практической конференции, посвященной Году науки и технологий в Российской Федерации. Курск, 2021. С. 326-328.

8. Прокопенко К.А., Савостьянова И.Л. Эконометрический анализ цен вторичного рынка квартир в Октябрьском районе г. Красноярска // Актуальные проблемы авиации и космонавтики. Сборник материалов VI Международной научно-практической конференции, посвященной Дню космонавтики. В 3-х томах. Под общей редакцией Ю.Ю. Логинова. Красноярск, 2020. С. 374-376.

9. Официальный сайт Территориальный орган Федеральной службы государственной статистики по Липецкой области. [Электронный ресурс] URL: <https://lipstat.gks.ru/> (дата обращения: 12.12.2022).

10. Яновский Л.П., Буховец А.Г. Введение в эконометрику: Учеб. пособие // Москва, 2017. (3-е издание, стереотипное). 265 с.

Сведения об авторе / Information about the author:

*Рязанцева Елена Анатольевна – к.ф.-м.н., доцент кафедры «Учет и информационные технологии в бизнесе», Липецкий филиал ФГОБУ ВО «Финансовый университет при Правительстве РФ», E-mail: EARYazantseva@fa.ru / **Ryazantseva Elena Anatolievna – Ph.D., Associate Professor of the Department of «Accounting and Information Technologies in Business», Lipetsk branch of the Federal State Budgetary Institution «Financial University under the Government of the Russian Federation», E-mail: EARYazantseva@fa.ru***
ORCID 0000-0002-7649-0566
SPIN-код: 7054-1215

Дата поступления статьи: 13.12.2022
Принято решение о публикации: 20.02.2023

Автор прочитал и одобрил окончательный вариант рукописи.
Конфликт интересов: автор заявляет об отсутствии конфликта интересов.