

Оценка стоимости коммерческой недвижимости в модели множественной регрессии с использованием RStudio.

Чори Владлен Александрович,
бакалавр,

Клочкова Ольга Ивановна

канд. физ-мат. наук, доцент кафедры математики и моделирования
Владивостокский государственный университет экономики и сервиса
Россия. Владивосток

E-mail: choryvlad@mail.ru; тел.: +79245295202

ул. Гоголя, 41, г. Владивосток, Приморский край, Россия, 690014

С помощью анализа базы данных существует возможность создавать прогноз цены на ближайшую перспективу. В статье приведен пример анализа базы данных интернет-ресурса «Farpost» - построены три модели на базе множественной регрессии.

Ключевые слова и словосочетания: анализ, модели, база данных, анализ базы данных, множественная регрессия, визуализация.

Valuation of commercial real estate in a multiple regression model using RStudio.

With the help of database analysis, it is possible to create a price forecast for the near future. The article provides an example of the analysis of the database of the Internet resource "Far post" - three models based on multiple regression are built.

Keywords: analysis, models, database, database analysis, multiple regression, visualization

Актуальность темы исследования обусловлено тем, что в условиях современной экономики необходимо просчитывать все шаги для достижения успеха организаций, и первым и самым важным шагом, является покупка, (аренда) помещения, в котором и организуется основной вид коммерческой деятельности.

Цель работы – анализ факторов влияющих на цену некоммерческих помещений на примере базы данных «Farpost».

Задачи работы:

- создать базу данных предприятий на продажу, выставленных на интернет-ресурсе «Farpost»;
- построить коэффициенты множественной регрессии при поддержке программы «RStudio»;
- визуализировать базу данных.

Для анализа была создана база данных на основе некоммерческих помещений, данные которых были взяты из интернет-ресурса «Farpost», куда входят такие переменные, как цена, площадь, район, вид помещения, исполнитель и состояние. В соответствии с этим, есть возможность построить модели, по которым сделать анализ и определить ключевые параметры, которые определяют цену: влияние на цену каждой переменной, например, расположение недвижимости и т.д. [1].

Множественная регрессия – это возможность продолжения линейной регрессии при помощи использования множества переменных, то есть использования более чем одного слагаемого в создании анализа [1]. Коэффициент детерминации R^2 показывает насколько тесной является связь между моделью и реальными данными. По значению коэффициента R^2 можно оценить какие факторы являются ключевыми при определении цены предприятия. Чем ближе R^2 к единице, тем точнее модель [2].

	A	B	C	D	E	F	G	H
1	Price	TypeOfPremises	RoomNumbering	District	DistrictNumbering	Square	Developer	Condition
2	15000000	Офис		1 Тихая	6	136.4	ООО "СЗ "Восточный ЛУЧ"	Сдан
3	54500000	Другое		4 Луговая	2	225.5	ООО СЗ "Альянс инвест"	Сдан
4	40000	Склад		3 Снеговая падь	3	209	ООО "СЗ "Восточный ЛУЧ"	Сдан
5	180000000	Другое		4 Луговая	2	1511	ООО СЗ "Альянс инвест"	Сдан
6	48500000	Другое		4 Луговая	2	257.5	ООО СЗ "Альянс инвест"	Сдан
7	45000000	Офис		1 Центр	0	198.5	Агентство Международный Центр Недвижимости	Сдан
8	12000000	Офис		1 Столетие	4	156.8	Частное лицо oooStandart	Сдан
9	20000000	Другое		4 Центр	0	142	ООО "ГК "Амира"	Сдан
10	7300000	Офис		1 Центр	0	91	Частное лицо Деловая Сеть	Сдан
11	11000000	Торговая точка		5 Центр	0	53.8	ООО "ГК "Амира"	Сдан
12	22873500	Торговая точка		5 Тихая	6	269.1	Fregatnovostroy	Сдан
13	2900000	Офис		1 Заря	8	40.5	Агентство ООО "КЛЭР"	Сдан
14	22000000	Другое		4 Сахарный ключ	10	450	Агентство АН "Метры"	Сдан
15	4650000	Офис		1 Центр	0	27	Агентство ООО Метрополис ВЛ	Сдан
16	12250000	Офис		1 Вторая речка	4	102.5	Агентство ИП Мусина С.Г.	Сдан
17	18900000	Офис		1 Вторая речка	4	70.9	Агентство ИП Мусина С.Г.	Сдан
18	122500000	Торговая точка		5 Вторая речка	4	513.7	Агентство ИП Мусина С.Г.	Сдан
19	134650000	Торговая точка		5 Вторая речка	4	616.2	Агентство ИП Мусина С.Г.	Сдан
20	58888888	Торговая точка		5 Океанская	11	505.8	Агентство ИП Мусина С.Г.	Сдан
21	10600000	Офис		1 Центр	0	38	Агентство ООО "Единый Ипотечный Центр"	Сдан
22	8300000	Офис		1 Вторая речка	4	50.6	Агентство ООО "Другое Дело"	Сдан
23	290000000	Офис		1 Вторая речка	4	3413.1	Частное лицо ОАО "Заря"	Сдан
24	3500000	Офис		1 Центр	0	28	Агентство ООО Центр недвижимости «АБРИС»	Сдан
25	62000000	Другое		4 Центр	0	330	Агентство Ваш риелтор Дмитрий Сергеев	Сдан
26	650000	Другое		4 Эгершельд	5	12	Частное лицо 1119756	Сдан
27	42800000	Склад		6 Пригород	9	2798	Частное лицо alexflash	Сдан
28	10900000	Офис		1 Некрасовская	1	348.3	Агентство ООО "Восточный город"	Сдан
29	7500000	Другое		4 Центр	0	26	Агентство ООО Центр недвижимости "АБРИС"	Сдан
30	26200000	Другое		4 Центр	0	171.8	Агентство Городской Риэлторский Центр	Сдан
31	35000000	Другое		4 Центр	0	274.9	Агентство ООО "ГК Амира"	Сдан
32	32606000	Другое		4 Центр	0	232.9	Агентство ООО "Восточный город"	Сдан
33	290000000	Офис		1 Вторая речка	4	3413.1	ОАО "Заря"	Сдан
34	105000000	Склад		6 Седанка	9	1041	Агентство проверенное	Сдан
35	7100000	Другое		4 Чуркин	7	62	Агентство "ТУНЭС"	Сдан
36	3100000	Другое		4 Горностай	12	124	Агентство ООО "Земельный брокер"	Сдан
37	10000000	Другое		4 Чуркин	7	83	Агентство ООО "Земельный брокер"	Сдан

Рис.1. Таблица базы данных интернет-ресурса «Farpost».

Таким образом, множественная регрессия, как математическая модель, помогает выявить значимые факторы, которые определяют цену продажи коммерческих помещений.

В дальнейшем, через анализ построенных моделей появляется возможность построить прогноз стоимости коммерческих помещений на ближайшую перспективу

Ниже составлены модели расчёта множественной регрессии:

- первая модель (model_0) включает зависимость логарифма цены от логарифма площади;
- во второй модели (model_1) добавляется еще слагаемое, связанное с районом города – их 14 - от центра Владивостока до Тихой;
- в третьей модели(model_2) к указанными в предыдущих пунктах добавляется слагаемое, детализирующее вид помещения (офис, торговая точка, склад и т.д.);
- в последней четвертой модели(model_3) кроме предыдущих появляется слагаемое, которое показывает зависимость цены от состояния сдачи объекта (сдано ли оно в эксплуатацию).

Расчёты, полученные в программе RStudio, указаны в таблице (Рис.2).

В таблице указаны коэффициенты перед слагаемыми в каждой модели. Видно, что перед слагаемыми, связанными с детализацией района (model_1) стоит минус, т.е. чем дальше от центра города, тем цена ниже.

Самыми дорогими районами по цене стали загородные элитные районы: 10 (Сахарный ключ), 11 (Океанская), и 12 (Горностай).

Если смотреть на соотношение цены к виду помещения, то можно сказать, что самыми дешевыми будут помещения с маленькой площадью (такие как кладовые и т.д.) с коэффициентом 0,088, а самыми дорогостоящими - складские помещения с коэффициентом 0,613.

```
R 4.1.2 · C:/Users/Moon Herim/Desktop/
model_3: lm(formula = log(Price) ~ log(Square) + DistrictNumbering + RoomNumbering + Developer:Condition, data = x)
```

	model_0	model_1	model_2	model_3
(Intercept)	12.190*** (0.177)	12.231*** (0.188)	12.215*** (0.165)	12.293*** (0.177)
log(Square)	0.856*** (0.035)	0.924*** (0.034)	0.931*** (0.029)	0.938*** (0.029)
DistrictNumbering: 1		-0.206 (0.148)	-0.212 (0.126)	-0.203 (0.126)
DistrictNumbering: 2		0.004 (0.188)	0.002 (0.158)	0.019 (0.157)
DistrictNumbering: 3		-1.433*** (0.237)	-0.921*** (0.212)	-0.970*** (0.212)
DistrictNumbering: 4		-0.365** (0.131)	-0.343** (0.111)	-0.340** (0.112)
DistrictNumbering: 5		-0.247 (0.200)	-0.188 (0.168)	-0.166 (0.168)
DistrictNumbering: 6		-0.597*** (0.169)	-0.585*** (0.143)	-0.578*** (0.143)
DistrictNumbering: 7		-0.677*** (0.200)	-0.626*** (0.169)	-0.616*** (0.168)
DistrictNumbering: 8		-0.290 (0.501)	-0.310 (0.421)	-0.332 (0.419)
DistrictNumbering: 9		-0.612* (0.301)	-0.489 (0.258)	-0.480 (0.257)
DistrictNumbering: 10		-0.968 (0.703)	-0.862 (0.590)	-0.785 (0.590)
DistrictNumbering: 11		-0.092 (0.703)	-0.316 (0.593)	-0.356 (0.591)
DistrictNumbering: 12		-1.737* (0.701)	-1.622** (0.589)	-1.674** (0.587)
DistrictNumbering: 13		-1.230*** (0.286)	-1.097*** (0.245)	-1.111*** (0.244)
DistrictNumbering: 14		-0.857** (0.282)	-0.740** (0.238)	-0.784** (0.240)
RoomNumbering: 3/1			-5.672*** (0.613)	-5.693*** (0.612)
RoomNumbering: 4/1			-0.134 (0.088)	-0.149 (0.088)
RoomNumbering: 5/1			0.196* (0.099)	0.204* (0.099)
RoomNumbering: 6/1			-0.560*** (0.173)	-0.585*** (0.172)
RoomNumbering: 7/1			-0.508 (0.272)	-0.527 (0.271)
Developer x Condition0				-0.046 (0.030)
Developer x Condition1				0.068 (0.100)
R-squared	0.691	0.758	0.834	0.836
N	267	267	267	267

Significance: *** = p < 0.001; ** = p < 0.01; * = p < 0.05

Рис.2. Таблица коэффициентов множественной регрессии построенных моделей

В зависимости от района наиболее выгодными для покупки являются Вторая речка и Садгород с коэффициентами 0,111 и 0,238 соответственно.

В третьей модели построены коэффициенты соотношения цены к сумме показателей площади, района, вида помещения, застройщика и состояния.

Всего 2 состояния – сдан (0), не сдан (1). Благодаря коэффициентам видно, что цена сданных помещений по отношению к не сданным оценивается, как 10/3, следовательно, сданные помещения практически в три раза дороже.

Сравнивая показатели в третьей модели, по сравнению со второй, можно сделать вывод, что учет застройщика и состояние помещения повышает параметр R^2 .

Таким образом, при увеличении количества слагаемых увеличивался коэффициент R^2 , то есть при построении первой модели (model_0) коэффициент $R^2=0,691$.

Во второй модели $R^2=0,758$.

В третьей модели этот параметр растет и максимального значения достигает в четвертой модели $R^2=0,836$

Исходя из вышеизложенного, можно сделать вывод, что при определении цены коммерческого предприятия были определены ключевые факторы.

Визуализация данных в RStudio – это инструмент, который помогает извлечь важную информацию из данных.

График соотношения цены и площади представлен на рис.3.

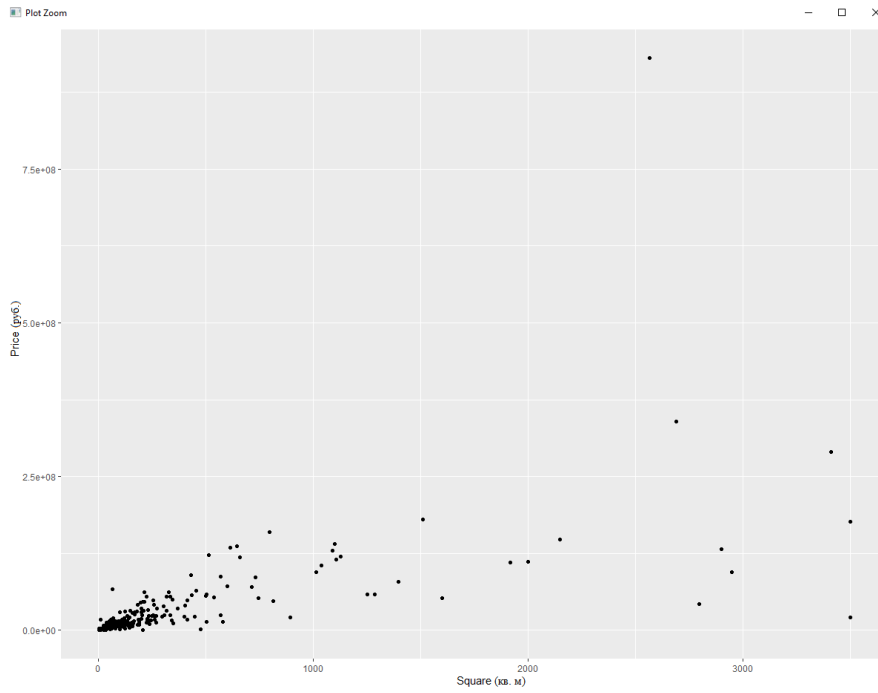


Рис.3. График соотношения цены и площади.

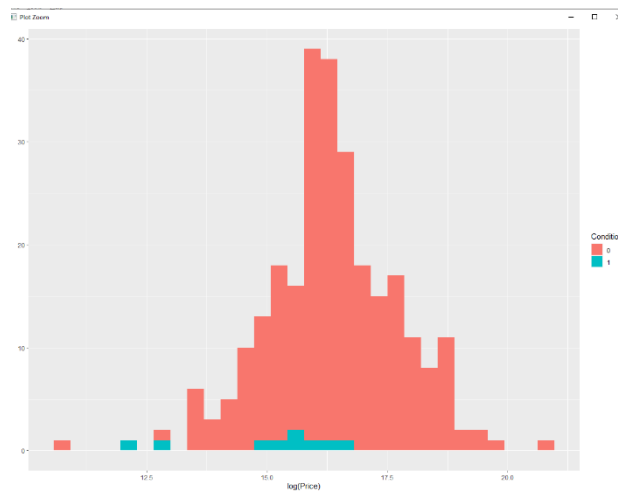


Рис.4. График соотношения цены и состояния сдачи объектов.

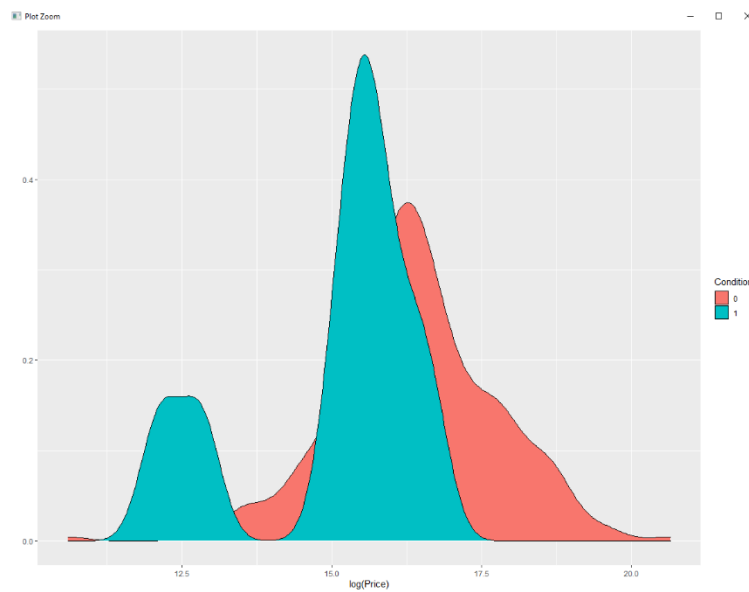


Рис.5. График плотности соотношения цены и состояния складских помещений.

На рис. 4 показана гистограмма помещений. Розовым цветом выделены сданные помещения, а зеленым помещения, документы которых оформленные не до конца. Видно, что сданных помещений смещена относительно гистограммы построенных, но до конца неоформленных помещений

График(рис.4) наглядно показывает соотношение сданных (розовый цвет)и не сданных объектов по мере роста логарифма цены и количества объектов. Видно, что наибольшее количество сданных и не сданных объектов находится в промежутке цены 15-17 млн рублей.

На примере складских помещений (рис.5)

На графиках (рис.5-6) показана плотность состояния сданных и не сданных складских объектов. Видно, что из небольшого количества не сданных помещений, большая часть находится в промежутке от 15 до 17 млн рублей, а в сданных помещениях график плавно возрастает от 12,5 млн рублей и убывает к 20 млн рублей(см. рис.6).

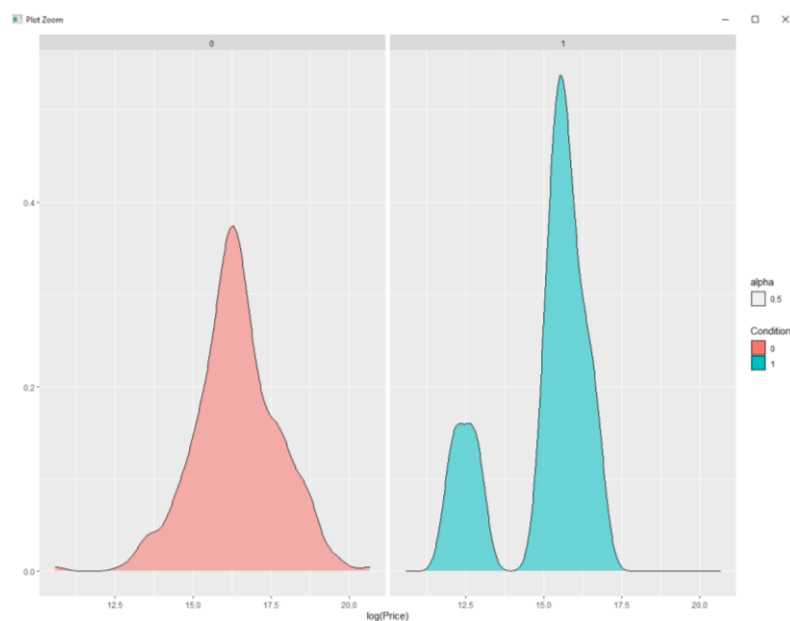


Рис.6. График плотности цены и состояния складских помещений в разрезе параметра «сдано» и «не сдано».

Таким образом, в ходе анализа построенных моделей множественной регрессии базы данных интернет-ресурса Farpost, были определены ключевые факторы, определяющие цену продажи коммерческой недвижимости.

Благодаря этому появляется возможность делать более точный прогноз цены в ближайшей перспективе.

1. R – множественная регрессия [Электронный ресурс]. – Режим доступа: <https://coderlessons.com/tutorials/kompiuternoe-programmirovaniie/learn-r-programmirovaniie/r-mnozhestvennaia-regressiia>, [Дата обращения – 15.04.2022]

2. R – значит регрессия/Хабр [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/post/350668/>, [Дата обращения – 15.04.2022]