

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Образовательная программа «Экономика»,
направление подготовки 38.03.01 «Экономика»

Бригинский Макар, Морозов Кирилл

ДОМАШНЯЯ РАБОТА 2 (ПРОЕКТ) ПО КУРСУ ЭКОНОМЕТРИКА-1

«Анализ месячной арендной квартирной платы в Москве за июнь 2023 года»

Преподаватель
Бывальцева-Станкевич
Анастасия Александровна

Москва, 2023 год

1. Введение

Цель нашего исследования - проанализировать влияние различных факторов на цену аренды квартиры. Для этого была собрана информация об объявлениях 1000 квартир с сайта <https://m2.ru/> за июнь 2023 года. Мы сосредоточим внимание на таких аспектах, как:

- 1) Price: стоимость аренды в месяц
- 2) Floor: этаж
- 3) Total area: Общая площадь квартиры
- 4) Rooms count: количество комнат
- 5) Has internet (дамми-переменная): наличие интернета
- 6) Has TV (дамми-переменная): наличие телевизора
- 7) Time metro: время до ближайшего метро
- 8) Floors house: количество этажей в доме

Главными задачами были выбраны построение логарифмически-линейной регрессии, оценка коэффициентов регрессии, а также проверка некоторых статистических гипотез, которые будут описаны далее.

Наше исследование имеет высокую актуальность, так как оно направлено на анализ влияния различных факторов на цены аренды квартир в современных условиях. С учетом роста значимости жилищного вопроса, особенно в контексте изменяющихся требований и стандартов, полученные результаты могут пролить свет на сложные взаимосвязи в жилищном рынке. Особое внимание будет уделено таким аспектам, как наличие интернета, телевизора, количество этажей и расстояние до общественного транспорта, что позволит выявить современные тенденции и влияние технологий на жилье.

Исходя из статьи "An econometric analysis of rent control", где исследуется влияние регулирования арендных ставок, мы видим, что регулирование может вызывать как положительные, так и отрицательные последствия. Это подчеркивает сложности в жилищном рынке, и исследование направлено на выявление подобных взаимосвязей в современном контексте.

В статье "Why have housing prices gone up?" поднимается вопрос о росте цен на жилье в США, в частности, о связи между новым строительством и ценами. Это может быть важным аспектом для нашего исследования, поскольку ограничения на новое строительство могут влиять на доступность жилья и, следовательно, на цены на аренду.

"The markets for real estate assets and space" предоставляет обзор методов непараметрической регрессии в контексте исследований в области жилища. Этот текст может быть полезен при выборе методов анализа данных в нашем исследовании.

Таким образом, наше исследование является актуальным, поскольку оно направлено на более глубокое понимание факторов, влияющих на цены аренды квартир, с учетом современных условий исследования.

В данных научных статьях использованные модели не применимы к тем результатам, которых мы хотим добиться, поскольку они учитывают влияние макроэкономических факторов и ситуации на других рынках (в статье 'Why have housing prices gone up?').

2. Экономическая модель

- 1) Floor: этаж (номер этажа, на котором расположена квартира)
- 2) Total area: Общая площадь квартиры (м²)

- 3) Rooms count: количество комнат (количество)
- 4) Has internet (дами-переменная): наличие интернета (1- присутствует; 0- отсутствует)
- 5) Has TV (дамми-переменная): наличие телевизора (1- присутствует; 0- отсутствует)
- 6) Time metro: время до ближайшего метро (минуты)
- 7) Floors house: количество этажей в доме (количество)

Данный набор переменных был выбран, поскольку это одни из самых распространённых критериев на сайтах по аренде/ продаже недвижимости. В силу этого нам стало интересно, все ли эти переменные имеют влияние на конечную цену аренды квартиры. Предположительно все переменные должны повышать цену арендным за исключением времени до метро, последнее действует обратным образом, поскольку будет затрачиваться больше времени у арендатора на дорогу.

В рамках нашего исследования нам интересно проверить следующие гипотезы:

- Квартиры, где интернет присутствует и где интернет отсутствует, нужно оценивать различными моделями.
- Убедиться в предположении, выдвинутом в прошлом абзаце, об уменьшении стоимости аренды в зависимости от увеличения времени до ближайшей станции метро.
- Мы предполагаем, что наличие телевизора в квартире влияет на стоимость аренды, больше, чем наличие интернета (сети WI-FI), поскольку первое является движимым имуществом, а интернет, в современном, мире присутствует у каждого в телефоне.

3. Предварительный анализ данных

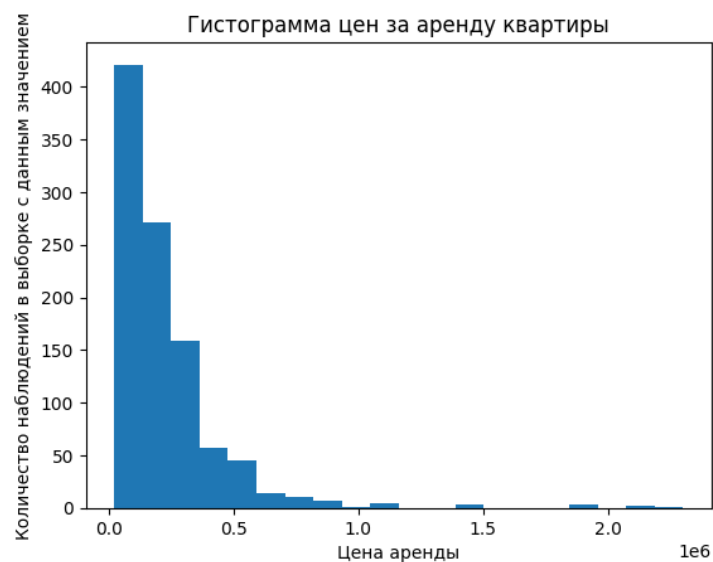
Для начала мы проанализировали независимость объясняющих переменных для более точной оценки. Для этого мы использовали оценочный коэффициент корреляции Пирсона и гипотезу о незначимости корреляции (расчёты предоставлены в файле Python).

	переменная 1	переменная 2	corr	P-value нулевой гипотезы	Принимается ли
0	floor	totalArea	0,020657	0,514096	ИСТИНА
1	floor	roomsCount	-0,03509	0,267661	ИСТИНА
2	floor	hasInternet	0,029257	0,355369	ИСТИНА
3	floor	hasTv	-0,0189	0,55056	ИСТИНА
4	floor	timeMetro	0,010834	0,732204	ИСТИНА
5	floor	floorsHouse	0,782086	3,4E-207	ЛОЖЬ
6	totalArea	roomsCount	0,857663	1,5E-290	ЛОЖЬ
7	totalArea	hasInternet	0,212201	1,21E-11	ЛОЖЬ
8	totalArea	hasTv	0,358895	9,16E-32	ЛОЖЬ
9	totalArea	timeMetro	-0,25145	6,93E-16	ЛОЖЬ
10	totalArea	floorsHouse	-0,08155	0,00988	ЛОЖЬ

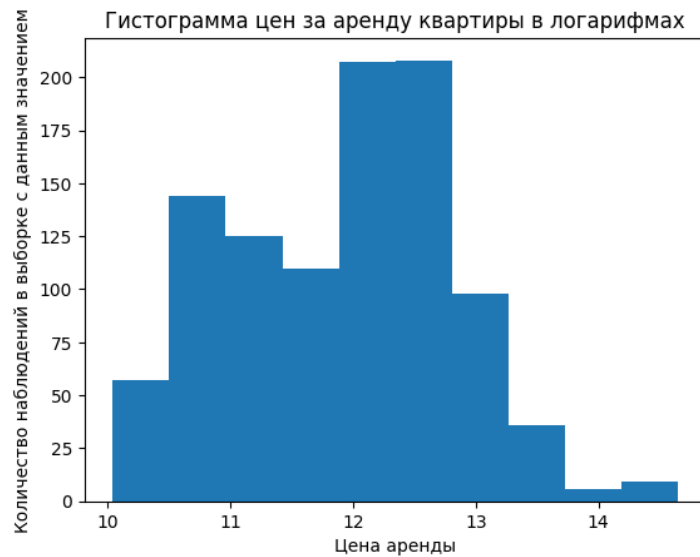
11	roomsCount	hasInternet	0,197047	3,28E-10	ЛОЖЬ
12	roomsCount	hasTv	0,391088	6,83E-38	ЛОЖЬ
13	roomsCount	timeMetro	-0,24916	1,29E-15	ЛОЖЬ
14	roomsCount	floorsHouse	-0,12156	0,000116	ЛОЖЬ
15	hasInternet	hasTv	0,273769	1,19E-18	ЛОЖЬ
16	hasInternet	timeMetro	-0,17812	1,42E-08	ЛОЖЬ
17	hasInternet	floorsHouse	-0,03542	0,263157	ИСТИНА
18	hasTv	timeMetro	-0,16008	3,61E-07	ЛОЖЬ
19	hasTv	floorsHouse	-0,04028	0,203131	ИСТИНА
20	timeMetro	floorsHouse	0,092119	0,00355	ЛОЖЬ

Хоть и не все гипотезы не отвергнуты, но оценочный коэффициент корреляции не слишком большой, что, по нашему предположению, не сильно исказит результаты.

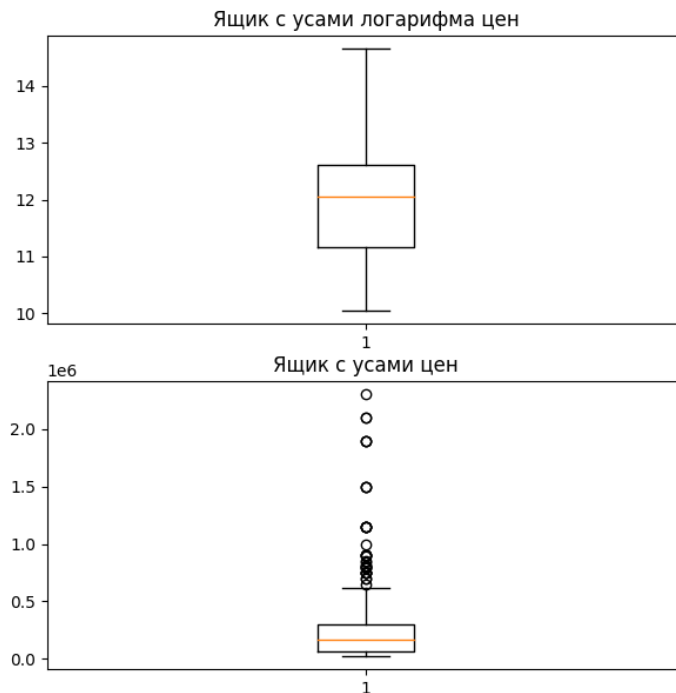
Далее мы решили проверить гистограмму целевой переменной, чтобы проверить наличие выбросов и смещений:



В силу слишком большого разброса мы решили логарифмировать целевую переменную:



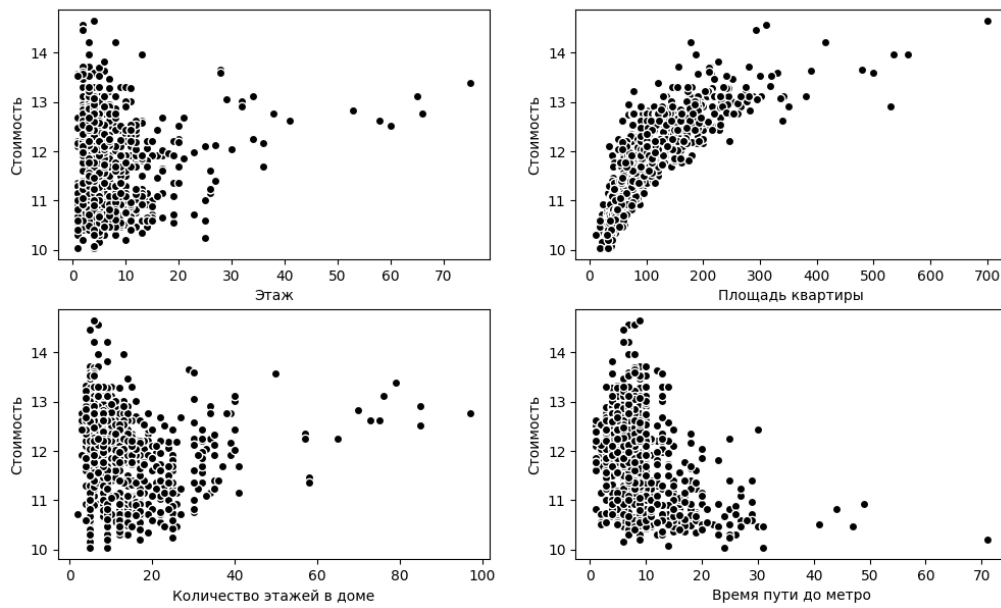
В силу этого разброс стал меньше и распределение стало похоже на нормальное. Диаграммы по типу ящик с усами также подтверждают данный вывод:



На этом анализ целевой переменной мы решили прекратить, поскольку после логарифмирования пропали все выбросы.

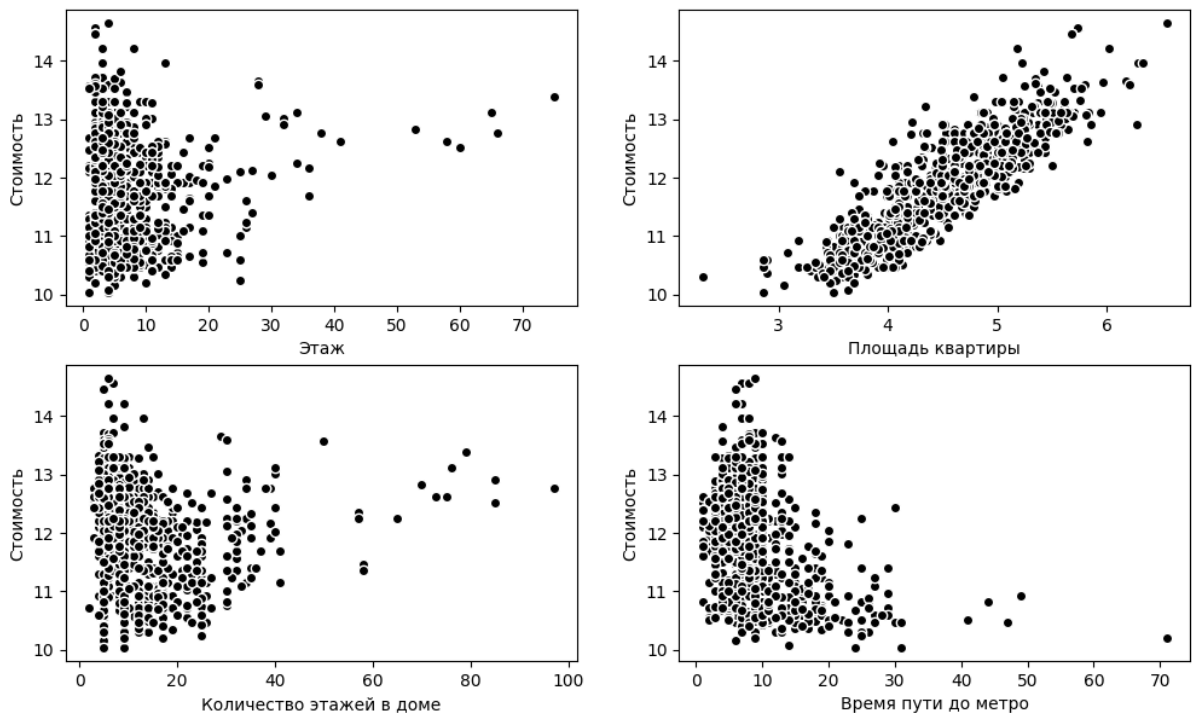
После этого мы решили посмотреть на связь объясняющих переменных и целевой переменной через диаграмму рассеивания. Были выбраны не все факторы, так как номер этажа и количество этажей являются дискретными переменными, и диаграмма рассеивания не поможет в установках связи между ними.

Графики рассеивания некоторых переменных



Явно видна положительная зависимость стоимости от площади квартиры, но она не линейна, поэтому далее мы будем использовать логарифм площади. Остальные параметры не показывают явной связи между целевой и объясняющей переменными.

Графики рассеивания некоторых переменных и логарифмов некоторых переменных



На одном из данных графиков (правый верхний) мы изобразили зависимость логарифма площади квартиры от логарифма стоимости, и мы можем наблюдать положительную линейную связь. На этом предварительный анализ данных завершен.

4. Оценка модели

После результата оценки модели мы получили следующие результаты:

	<i>Dependent variable:</i>
	Log (price)
floor	-0.001 (0.003)
Log (totalArea)	1.218*** (0.046)
roomsCount	-0.056*** (0.021)
hasInternet	0.017 (0.039)
hasTv	0.223*** (0.032)
timeMetro	-0.016*** (0.002)
floorsHouse	0.0002 (0.002)
Constant	6.523*** (0.160)
Observations	1,000
R ²	0.802
Adjusted R ²	0.801
Residual Std. Error	0.404 (df = 992)
F Statistic	574.430*** (df = 7; 992)
Note:	* ** *** p<0.01

Исходя из оценки сразу можно заметить, что наша гипотеза об отрицательном влиянии времени от квартиры до метро подтвердилась, поскольку коэффициент статистически значим.

Далее мы приступили к оценке гипотезы о разных генеральных совокупностях (гипотеза о том, что квартиры с интернетом и без интернета взяты из разных генеральных совокупностей). Она была выбрана, потому что почти в каждой квартире есть интернет и его отсутствие может вызвать подозрение о нестандартности квартиры. Все расчёты приведены в Python. P-value оказался равен 0.0002, гипотеза отвергается, что означает, что выборки неоднородны.

После этого мы проверили гипотезу, что наличие телевизора влияет на стоимость больше, чем наличие интернета. Проверка была выполнена при помощи односторонней проверки общей линейной гипотезы и были получены следующие результаты.

Statistic	N	Mean	St. Dev.	Min	Max
Res.Df	2	992.500	0.707	992	993
RSS	2	163.043	1.641	161.883	164.203
Df	1	1.000		1	1
Sum of Sq	1	2.320		2.320	2.320
F	1	14.219		14.219	14.219
Pr(> F)	1	0.0002		0.0002	0.0002

Как мы можем наблюдать правосторонняя альтернатива равна 0.0002 (подробные расчёты приведены в файле R), что означает, что гипотеза о равенстве отвергнута, в пользу правосторонней альтернативы. На этом оценка модели закончена.

5. Выводы

После проделанной нами работы мы выяснили, что:

- изменение площади квартиры на 1% изменит стоимость квартиры на 1,22% при прочих равных
- этаж статистически незначим
- количество комнат отрицательно влияет на стоимость при прочих равных на любом уровне значимости
- наличие не оказывает статистического влияния
- наличие телевизора добавляет 0.22% к цене квартиры
- увеличение времени до метро отрицательно влияет на цену (каждая минута уменьшает стоимость на 0.02%).
- количество этажей в доме статистически незначимо

Что касемо гипотез, подтвердилось наше предположение о том, что квартиры в самом деле из различных генеральных совокупностей, это скорее всего, в самом деле, вызвано тем, что отсутствие сети WI-FI в квартире является нестандартной ситуацией и они оцениваются по-своему.

Гипотеза об отрицательном влиянии метро, подтверждается, поскольку арендаторы не хотят тратить много времени на дорогу.

Гипотеза о 'превосходстве' телевизора подтверждается, поскольку телевизор является сегодня многофункциональной вещью (монитор под игровую приставку), а интернет, итак, имеется у каждого в смартфоне.

Научная литература:

https://www.jstor.org/stable/1830211?read-now=1&seq=1#page_scan_tab_contents

https://www.nber.org/system/files/working_papers/w11129/w11129.pdf

https://www.researchgate.net/publication/287232776_A_review_of_hedonic_pricing_models_in_housing_research