­­­ Федеральное государственное образовательное бюджетное учреждение высшего образования

**«Финансовый университет при Правительстве Российской Федерации»**

**(Финансовый университет)**

Факультет «Информационных технологий и анализа больших данных»

Направление подготовки

01.03.02 Прикладная математика и информатика

Анализ данных и принятие решений в экономике и финансах

***Индивидуальное домашнее задание***

По дисциплине «Эконометрика» на тему:

«Эконометрическое моделирование стоимости квартир в г. Реутов»

Выполнил студент

2 курса, учебной группы ПМ21-3

Меньшиков Илья Владимирович

Москва 2023

Эконометрическое моделирование стоимости квартир в г. Реутов

**1. Постановка задачи**

Разработать модель, которая позволит прогнозировать цены на квартиры в г. Реутов в зависимости от факторов, влияющих на эту стоимость.

**2. Описание используемых данных**

**2.1. Источник данных**

Поиск готовых, качественных и актуальных данных оказался сложным. Поэтому данные были получены с сайта по продаже недвижимости cian.ru. Для этого был использован проект, найденный на github.com. Этот проект позволяет указать нужные параметры и сохранить данные в формате csv.

**2.2. Описание данных**

Было решено использовать объявления о продаже квартир в городе Реутов.

Реутов - это наукоград, расположенный в Московской области России. Он находится в 20 километрах к востоку от Москвы и является частью Московской агломерации. В Реутове расположены множество научных институтов, включая Институт ядерной физики им. Г. И. Будкера, Институт микроэлектроники технического университета, Институт радиотехники и электроники РАН и другие. Также здесь находится много предприятий, занимающихся разработкой и производством электроники, приборов и других технологических изделий.

После парсинга датасет содержит следующие факторы:

- district - район

- underground - метро

- street - улица

- floor - этаж

- floors\_count - общее количество этажей

- total\_meters - общая площадь

- living\_meters - жилая площади

- kitchen\_meters - площадь кухни

- rooms\_count - количество комнат

- year\_construction - год постройки здания

- price\_per\_month - стоимость в месяц

- price\_per\_m2 - стоимость на квадратный метр

- commissions - комиссия, взымаемая при заселении

- author - автор объявления

- author\_type - тип автора

- phone - номер телефона в объявлении

- link - ссылка на объявление

Часть из них являются ненужными для дальнейшей работы, поэтому их следует исключить.

Зависимая переменная: Y - оценка стоимости квартиры.

После преобразований остались следующие факторы:

- Количество комнат

- Год постройки

- Общая площадь

- Жилая площадь

- Площадь кухни

- Этаж

- Всего этажей

- Автор объявления (0 - агентство недвижимости, риелтор, 1 - владелец)

**3. Расчет параметров регрессионного уравнения**

Была построена матрица парных коэффициентов корреляции для оценки значимости этих факторов.

- Y - Цена

- X1 - Автор объявления

- X2 - Этаж

- X3 - Всего этажей

- X4 - Число комнат

- X5 - Общая площадь

- X6 - Год постройки

- X7 - Жилая площадь

- X8 - Площадь кухни

Таблица 1 Матрица парных коэффициентов корреляции

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Желтым цветом выделил сильную корреляционную связь.

Анализ матрицы парных коэффициентов корреляции показывает, что:

• корреляционная зависимость стоимости квартиры и факторов «Общая площадь» и «Жилая площадь» прямая, сильная – с большей общей и жилой площадью цена на квартиры выше;

• корреляционная зависимость стоимости квартиры и фактора «Автор объявления» прямая, очень слабая, т.е. автор объявления в небольшой степени влияет на стоимость квартиры;

• корреляционная зависимость стоимости квартиры и фактора «Этаж» прямая, слабая: чем выше этаж квартиры, тем немного выше стоимость квартиры;

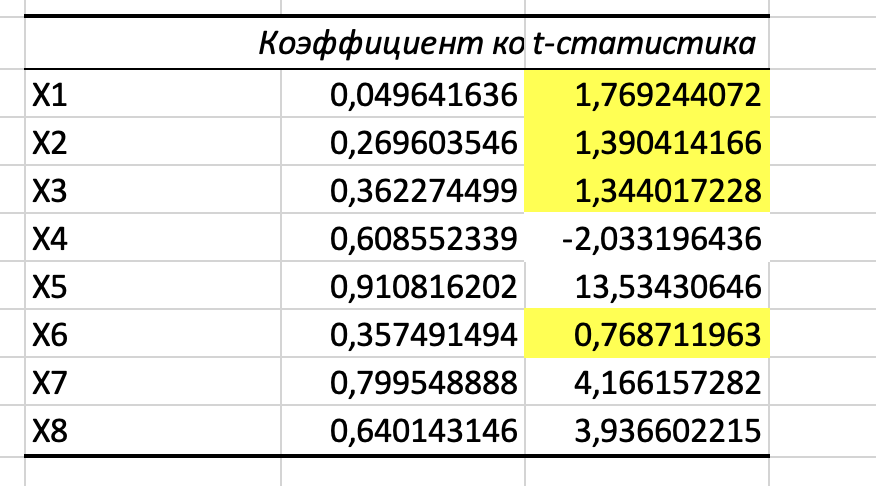
• корреляционная зависимость стоимости квартиры и факторов «Всего этажей» и «Год постройки» прямая, умеренная: чем больше этажей в здании и более новый дом, тем стоимость выше.

• корреляционная зависимость стоимости квартиры и факторов «Число комнат» и «Площадь кухни» прямая, средняя: чем больше комнат, а также, чем больше площадь кухни, тем стоимость квартиры выше.

Чтобы проверить значимость коэффициентов корреляции я использую t-критерий Стьюдента. Табличное значение t-статистики Стьюдента с уровнем значимости 0,05.

t-табл = 1,96347

Таблица 2



Факторы X4, Х5, Х7, Х8 признаются статистически значимыми, что свидетельствуют о наличии устойчивой линейной связи. Так как |t-расч| > t-табл

Проверка модели на мультиколлиниарность.

Таблица 3 Парных коэффициентов корреляции X1-X8

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Связи между объясняющими переменными достаточно тесная, особенно между переменными X3 и X6, X4 и X5, X4 и X7, X5 и X7, X5 и X8, что указывает на интеркорреляцию этих переменных. Для проверки обоснованности включения этих переменных в модель использую правило из учебника И. И. Елисеевой

Формула 1 из учебника И.И. Елисеева стр. 99

Изображение выглядит как текст, рукописный текст, Шрифт, белый

Автоматически созданное описание

X3 и X6, X4 и X5, X4 и X7 правило не выполняется

X5 и X8 правило выполняется

Далее провожу VIF тест

Таблица 4 VIF тест

Изображение выглядит как текст, Шрифт, снимок экрана, число

Автоматически созданное описание

Все значения превышают 3, что указывает на наличие мультиколлинеарности.

Из-за наличия сильной карреляции между факторами, не выполнения правила, а также результата VIF теста исключаю X7, X6, X4, X3.

Таблица 5 Парных коэффициентов корреляции Y, X1, X2, X5, X8

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

X5 и X8 правило выполняется (0,91 > 0,63 и 0,64 > 0,63)

Таблица 6 Y(X1, X2, X5, X8) Регрессионный анализ

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

P-значение X1 0,2628 > 0.05, а также t-набл < t-табл (1,12 < 1,96). Следовательно фактор X1 не является значимым и его можно исключить.

Таблица 8 Y(X2, X5, X8) Регрессионный анализ

Изображение выглядит как текст, Шрифт, снимок экрана, число

Автоматически созданное описание

t-табл = 1,963448464

F-табл = 2,61796338

F-набл = 1173,20505

F-набл > F-табл, таким образом расхождение между вычисленными дисперсиями существенна и носит неслучайный характер.

Ŷ = 292347,04 + 28884,68\*X2 + 177414,61\*X5 + 84360,29\*X8

**4. Анализ адекватности регрессионного уравнения**

**4.1. Показатели качества подгонки**

Регрессионное уравнение имеет высокий показатель множественного R, что говорит о хорошей подгонке модели к данным. Коэффициент детерминации R^2=0,83768147 высокий, что говорит о том, что модель объясняет большую долю изменчивости зависимой переменной.

Дисперсионный анализ показывает, что регрессия статистически значима (F-набл > F-табл), а коэффициенты значимы (P-значение < 0,05) при всех трех факторах (X2, X5, X8).

Таким образом, данное уравнение регрессии адекватно и может быть использовано для прогнозирования зависимой переменной на основе значений факторов.

**4.2. Проверка условий для получения «хороших» оценок МНК**

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Рисунок 1 Графики остатков для X2, X5, X8

Для проверки на гетероскедастичность использовал тест Голдфелда-Квандта.

H0: присутствует гомоскедастичность.

H1: присутствует гетероскедастичность.

Для X2: F-набл = 2,390638381

Для X5: F-набл = 6,219502607

Для X8: F-набл = 4,59190165

F-крит = 1,245800072

F-набл для всех факторов > F-крит, следовательно нулевая гипотеза отвергается, гетероскедастичность есть.

**4.3. Экономический смысл коэффициентов регрессии**

Коэффициенты эластичности позволяют определить силу влияния каждого фактора на цену квартиры. Полученные значения эластичности для каждого фактора составляют:

Э2 = 0,02375095

Э5 = 0,86893236

Э8 = 0,08219391

Для общей площади (X5): 0,8689. При увеличении общей площади на 1%, стоимость квартиры увеличивается на 0,8689%.

Для этажа (X2): 0,2375. То есть, при увеличении значения этажа на 1%, стоимость квартиры возрастает на 0,2375%.

Для площади кухни (X8): 0,08219. При увеличении площади кухни на 1%, стоимость квартиры увеличивается на 0,08219%.

Рассчитаем относительную ошибку аппроксимации по формуле:



Средняя ошибка аппроксимации равна 11,03482718. Среднее абсолютное отклонение между предсказанными и фактическими значениями составляет 11,03482718.

Коэффициент регрессии X2 показывает, что при увеличении этажа стоимость квартиры увеличивается на 28884,68

Коэффициент регрессии X5 показывает, что при увеличении общей площади квартиры на 1кв. м ее стоимость увеличивается на 177414,61

Коэффициент регрессии X8 показывает, что при увеличении площади кухни на 1кв. м. стоимость квартиры увеличивается на 84360,29

**5. Выводы**

В результате проведенной работы была построена модель для прогнозирования цен на квартиры в г. Реутов. Были определены значимые факторы, которые влияют на цену квартиры, и решена проблема мультиколлиниарности. Также была произведена проверка на гетероскедастичность. На основе выбранных факторов было построено регрессионное уравнение, которое может использоваться для прогнозирования цены квартиры. Был проведен анализ адекватности уравнения и проверен экономический смысл его коэффициентов.

**Используемые источники:**

1. <https://github.com/lenarsaitov/cianparser?ysclid=lhrqb7xtpf820326163>. Данный репозиторий предоставляет инструменты и возможности для получения данных с сайта.

2. cian.ru. Это один из ведущих онлайн-порталов по поиску и аренде недвижимости в России.

3. Учебник "Эконометрика" И. И. Елисеева, 2012 г.