

Анализ цен на недвижимость города  
Темиртау(Казахстан)  
по предмету «Статистическое оценивание и  
планирование наблюдений».  
Преподаватель: Булгакова Мария Александровна  
Выполнил: Янн Александр, группа 19.Б02-ПУ

Пункт 1.

В этой работе проанализированы цены на  
недвижимость в городе Темиртау (Казахстан).  
В качестве объясняющих переменных были выбраны  
такие параметры:

- 1.Количество комнат(1,2,3,4)
- 2.Площадь квартиры ( $m^2$ )
- 3.Район (в черте города или за городом) (0,1)
- 4.Год постройки(гггг)
- 5.Площадь кухни( $m^2$ )
- 6.Этаж(n)
- 7.Санузел (совмещенный или раздельный) (0,1)

Объясняемая переменная **цена**

Количество наблюдений 50. Взяты с сайта krisha.kz  
Данные параметры являются основными и влияют на  
ценообразование квартир, так как от них зависит  
комфортность условий проживания.

Переменная **Район** является фиктивной и была  
обработана, как **1** центре города, **0** на окраине.

Переменная **Санузел**, также является фиктивной и  
была обработана, как **0** совмещенный, **1** раздельный.

Пример первых десяти данных:

1	Комнат(x1)	Площадь(x2)	Район(x3)	Год постройки(x4)	Площадь кухни(x5)	Этаж(x6)	Санузел(x7)	Цена(y)
2	1	31,3	0	1970	6	5	1	5700000
3	3	55,3	0	1977	5,9	8	1	10000000
4	2	50	1	1967	6	4	0	13000000
5	4	80	1	1974	12	9	1	19500000
6	2	41	1	1970	5	2	1	12000000
7	3	47	0	1966	6	2	1	10200000
8	4	80	1	1980	10	1	1	25000000
9	1	31	0	1980	6	5	0	7500000
10	3	64,2	1	1991	8,9	5	1	20500000

## Пункт 2.

Проверка на мультиколлинеарность.

Составим корреляционную матрицу Q.

	Комнат(х1)	Площадь(х2)	Район(х3)	Год постройки(х4)	Площадь кухни(х5)	Этаж(х6)	Санузел(х7)	Цена(у)
Комнат(х1)	1	0,823046628	0,480944531	0,028406358	0,199385978	0,021414929	0,233663824	0,617236074
Площадь(х2)	0,823046628	1	0,522578294	0,357884879	0,580966399	0,047937767	0,314181148	0,842782473
Район(х3)	0,480944531	0,522578294	1	0,354182055	0,291690352	0,119996	0,21821789	0,68957134
Год постройки(х4)	0,028406358	0,357884879	0,354182055	1	0,610554526	0,064203078	0,330290321	0,452253706
Площадь кухни(х5)	0,199385978	0,580966399	0,291690352	0,610554526	1	-0,015978389	0,131137605	0,637265637
Этаж(х6)	0,021414929	0,047937767	0,119996	0,064203078	-0,015978389	1	0,237122203	0,021545598
Санузел(х7)	0,233663824	0,314181148	0,21821789	0,330290321	0,131137605	0,237122203	1	0,266808979
Цена(у)	0,617236074	0,842782473	0,68957134	0,452253706	0,637265637	0,021545598	0,266808979	1

Определитель матрицы Q = 0,038045414. Стремится к нулю, дополнительно, проведем следующий тест.

Вычислим  $\text{Det}(X^T X) = 0$ . Из проведенных тестов, мы можем утверждать о мультиколлинеарности.

## Пункт 3.

Из модели уберем те факторы которые имеют большую корреляцию друг с другом ( $\geq 0,7$ ). В нашем случае это

**Площадь** и количество **Комнат**. Удалим количество **Комнат**. Составим новую корреляционную матрицу Q.

	Площадь(х2)	Район(х3)	Год постройки(х4)	Площадь кухни(х5)	Этаж(х6)	Санузел(х7)	Цена(у)
Площадь(х2)	1	0,52257829	0,357884879	0,580966399	0,047937767	0,314181148	0,842782473
Район(х3)	0,52257829	1	0,354182055	0,291690352	0,119996	0,21821789	0,68957134
Год построй	0,35788488	0,35418205	1	0,610554526	0,064203078	0,330290321	0,452253706
Площадь ку	0,5809664	0,29169035	0,610554526	1	-0,015978389	0,131137605	0,637265637
Этаж(х6)	0,04793777	0,119996	0,064203078	-0,015978389	1	0,237122203	0,021545598
Санузел(х7)	0,31418115	0,21821789	0,330290321	0,131137605	0,237122203	1	0,266808979
Цена(у)	0,84278247	0,68957134	0,452253706	0,637265637	0,021545598	0,266808979	1

Определитель матрицы Q = 0,209666424; Не стремится к нулю, проверим дальше.

Вычислим  $\text{Det}(X^T X) = 6,25426E+15$ . Не стремится к нулю.

Проведем vif-test:

х2=у	r^2(1)	0,5242275	vif1	2,1018448	мультиколлинеарности нет
х3=у	r^2(2)	0,3285022	vif2	1,4892081	vif<5
х4=у	r^2(3)	0,4771901	vif3	1,9127411	
х5=у	r^2(4)	0,5619261	vif4	2,2827197	
х6=у	r^2(5)	0,0676449	vif5	1,0725528	
	r^2(6)	0,2393556	vif6	1,3146748	

Итог: мультиколлинеарности нет.

## Пункт 4.

	Коэффициенты	ндртная ошт-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
У-пересечение	-2324004,456	62708429,87	-0,03706048	0,9706083	-128787605,8	124139596,9	124139596,9
Переменная X 1	160818,5359	27558,11256	5,835615031	6,338E-07	105242,3052	105242,3052	216394,7665
Переменная X 2	3069876,967	666427,0171	4,606471359	3,631E-05	1725898,8	4413855,134	4413855,134
Переменная X 3	1109,168368	32036,3532	0,034622179	0,9725412	-63498,29523	65716,63197	65716,63197
Переменная X 4	431555,7414	181182,4762	2,381884554	0,0217195	66166,45504	796945,0278	796945,0278
Переменная X 5	-76007,4883	113109,6752	-0,67198043	0,5051908	-304114,8879	-304114,8879	152099,9113
Переменная X 6	38135,7089	683194,6079	0,055819687	0,955744	-1339657,527	1415928,945	1415928,945

Можно обратить внимание, что при фиксации всех остальных параметров, кроме X5(площадь кухни), чем больше будет площадь кухни, а следовательно жилая площадь будет меньше, цена на квартиру будет меньше.

Пункт 5.

Проверим значимость регрессии в целом. Вычислим значение F-статистики и ее критическую точку.

F	35,55095773
F <sub>расп обр</sub>	2,318498031

Можем сделать вывод о том, что регрессия в целом значима. Рассчитаем значение t-статистики для каждого коэффициента.

1	5,835615031
2	4,606471359
3	0,034622179
4	2,381884554
5	-0,671980431
6	0,055819687
t	2,016692199

Значимыми переменными являются **Площадь, Район, Площадь кухни.**

Пункт 6.

Удалим такие переменные: **Этаж, год постройки, санузел.** Построим регрессионную модель.

	Коэффициенты
У-пересечение	-432099,0194
Площадь X 1	160867,1329
Район X 2	3029464,481
Площадь кухни X 3	439812,0182

Пункт 7.

Добавим фактор **площадь.**

Нормиров

анный R- 0,7042

квадрат 4651

Добавим в модель.

К фактору **площади** добавим **район.**

Нормиров

анный R- 0,7869

квадрат 8534

Порог в 0,1 не преодолен -> не добавляем в модель.

К фактору **площади** добавим фактор **год постройки**

Нормирова 0,72508

нный R-  
квадрат 526

Порог в 0,1 не преодолен -> не добавляем в модель.

К фактору **площади** добавим фактор **площадь кухни**.

Нормирова  
нный R- 0,7322  
квадрат 5586

Порог в 0,1 не преодолен -> не добавляем в модель.

К фактору **площади** добавим фактор **этаж**

Нормирова  
нный R- 0,6983  
квадрат 254

Порог в 0,1 не преодолен -> не добавляем в модель.

К фактору **площади** добавим фактор **санузел**

Нормирова  
нный R- 0,6979  
квадрат 5862

Порог в 0,1 не преодолен -> не добавляем в модель.

Пункт 8.

Проанализируем полученные модели по коэффициенту  
нормированному R-квадрату

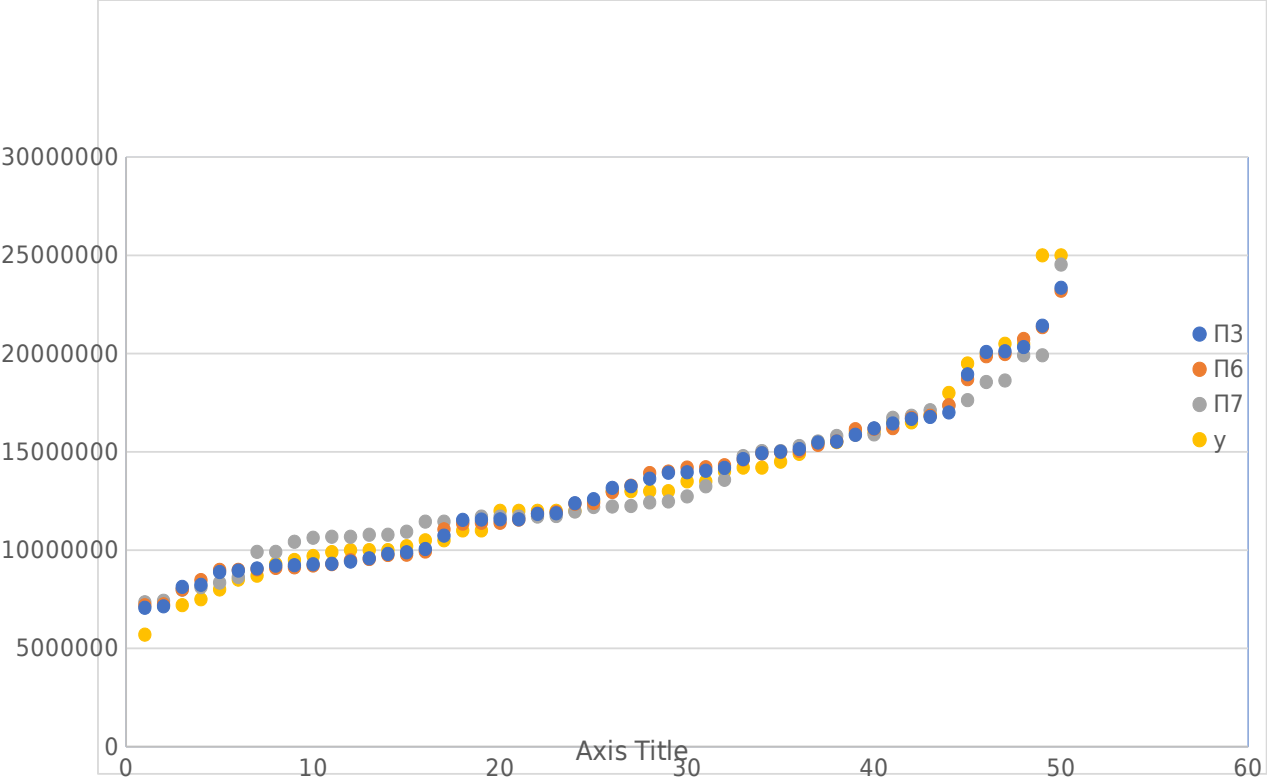
В первой модели от 6 переменных значение нормированного  
R-квадрата = 0,808822077978169.

Рассмотрим модель полученную в Пункте 6 от трех  
переменных. Нормированный R-квадрат =  
0,819378119718831. Значение выросло, из этого следует, что  
некоторые факторы превносили «шумы» в модель.

Рассмотрим модель полученную в Пункте 7 из одной  
переменной. Нормированный R-квадрат =

0,704246511036873. Значение упало на ~ 0,1. Если  
необходима «простая» оценка можно использовать только  
общую площадь.

Пункт 9.  
График цен(сортированных) от n



Пункт 10.

	Доверительный интервал для параметра	
	Левая граница	Правая граница
Площадь	105242,3052	216394,7665
Район	1725898,8	4413855,134
Площадь кухни	66166,45504	796945,0278

Пункт 11.

	bi(стандартизированные)	
Площадь	0,528614421	
Район	0,346506823	
Площадь кухни	0,229085723	

Чем больше стандартизированный коэффициент, тем больший вклад вносит в итоговую цену квартиры данный параметр. В нашем примере, самым значимым является площадь.

#### Пункт 12.

Коэффициент детерминации показывает долю объясненности дисперсии зависимой переменной дисперсиями рассматриваемых в модели факторов.

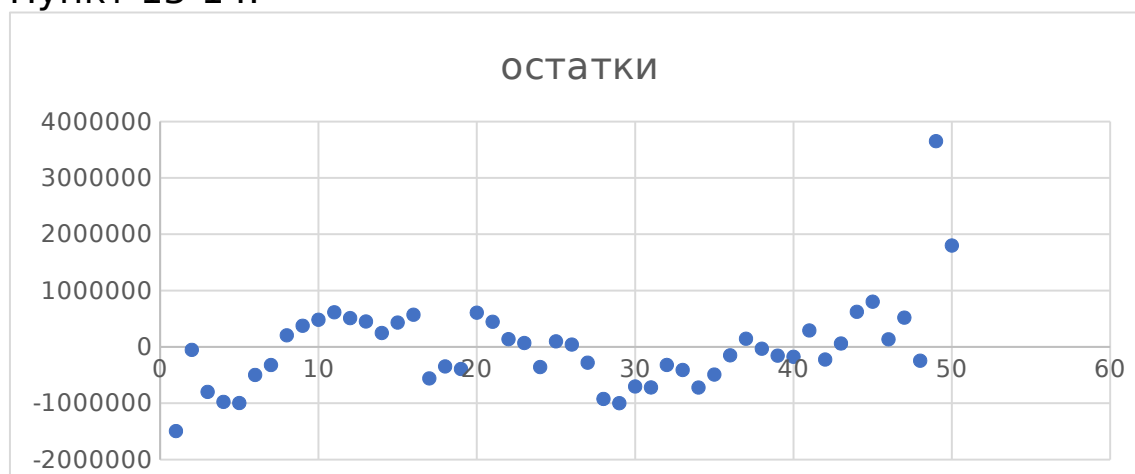
R-квадрат	0,830436602
Нормированный R-квадрат	0,81937812

Значение 0,8 является хорошим результатом.

Средняя ошибка аппроксимации данной модели равна ~10%, а среднее абсолютное отклонение равняется  $MAD = 1288977,0$ .

Первое значение показывает, на сколько процентов в среднем модель ошибается относительно известных значений, а второе показывает, на сколько в среднем модельная цена отличается от реальной.

#### Пункт 13-14.



Заметим, что в конце присутствует несколько выбросов, в общем и целом дисперсия кажется постоянной. Есть основания ожидать автокорреляцию. Можно ожидать нулевое мат. ожидание.

#### Тест на нулевое мат. ожидание

Статистика

-5,91415E-16

Значение крит. точки -1,96.

В критическую область не попадает, из этого следует принятие нулевой гипотезы о равенстве математического ожидания нулю.

## Тест на гомоскедастичность

Goldfeld-Quandt test

data: m2

GQ = 4.7479, df1 = 16, df2 = 16, p-value = 0.001682

alternative hypothesis: variance increases from segment 1 to 2

Воспользовавшись тестом Гольфельда-Квандта принимается альтернативная гипотеза, о том что гомоскедастичность отсутствует.

Возможная причина этого – выбросы, которые можно увидеть на графике.

## Тест на автокорреляцию

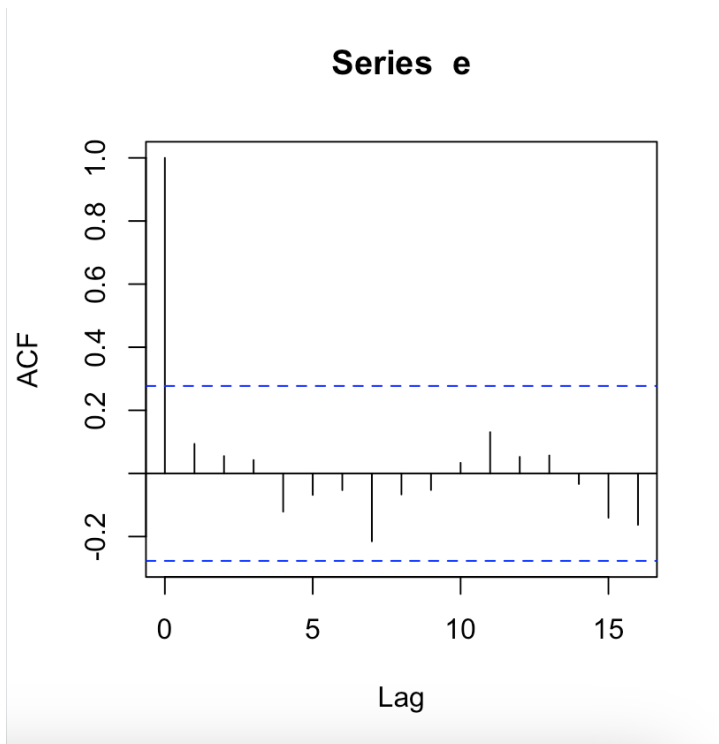
Durbin-Watson test

data: model

DW = 1.7938, p-value = 0.2294

alternative hypothesis: true autocorrelation is greater than 0

Автокорреляции первого порядка не обнаружена. Проверим на  $k > 2$



Автокорреляции высших порядков также не обнаружено.

### **Тест на нормальное распределение.**

Тест Шапиро-Уилка

Shapiro-Wilk normality test

```
data: e
W = 0.96153, p-value = 0.1031
```

$p\text{-value} > 0,05$  из этого следует, что нет оснований отрицать что остатки распределены нормально.

Пункт 15.

Прогноз для значений.

Площадь ь	Район	Площадь кухни	Прогнозируемая цена
77,9346	0	10,3849	16672420,46
77,9346	1	10,3849	19701884,94

Пункт 16.



Доверительный интервал для точечного прогноза 0его квартала		
s ост	1838788,747	
m1y^p	2021127,709	
tkp	2,012895599	
12604101,39 yp		20740739,53
Доверительный интервал для точечного прогноза 1его квартала		
s ост	1838788,747	
m2y^p	1915541,298	
15846100,29 yp		23557669,59
Доверительный интервал для мат ожидания прогноза 0его квартала		
s ост	1838788,747	
mMx(y)	838935,7304	
14983730,42		18361110,5
Доверительный интервал для мат ожидания прогноза 1его квартала		
s ост	1838788,747	
mMx(y)	536800,1562	
18621362,27		20782407,61