ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук Департамент программной инженерии

Дисциплина: Прикладные методы математической статистики

Моделирование цены двухкомнатной квартиры в Москве на основании регрессионного анализа

Исполнители:

студенты группы БПИ203

Дмитриева В. А.

Насыхова А. А.

Описание данных

Площадь (M^2) – общая площадь квартиры;

Площадь кухни (M^2) – площадь кухни в квартире;

Жилая площадь $(м^2)$ – общая площадь жилых комнат в квартире;

Год постройки – год постройки здания, в котором расположена квартира;

Расстояние до метро (м) — расстояние от здания, в котором расположена квартира, до ближайшей станции метро;

Вид объекта — переменная, значение которой равно 0, если квартира продается на вторичном рынке, или 1, если в новостройке;

Балкон/**Лоджия** — переменная, значение которой равно 0, если в квартире нет балкона/лоджии, или 1 в противном случае;

Лифт – переменная, значение которой равно 0, если в здании, в котором расположена квартира, нет лифта, или 1 в противном случае;

Парковка — переменная, значение которой равно 0, если в здании/жилом комплексе, в котором расположена квартира, нет парковки, или 1 в противном случае;

Ремонт – переменная, значение которой равно 0, если квартира не нуждается в ремонте, или 1 в противном случае;

МКАД – переменная, значение которой равно 0, если здание, в котором расположена квартира, находится за пределами МКАД, или 1 в противном случае;

Цена (руб.) – цена за квартиру;

Первый/последний — переменная, значение которой равно 0, если квартира расположена не на первом и не на последнем этаже здания, или 1 в противном случае;

Описательная статистика

цена		т об постіройки		<i>Расстояние оо метгро</i>		Площаоь	
0	40704277 70	0	2000 24400	0	744 0747040	0	FC 000F70CF
Среднее	18764377,78		2008,34188		744,8717949		56,22957265
Стандартная ошибка		Стандартная ошибка		Стандартная ошибка		Стандартная ошибка	1,415834906
Медиана		Медиана		Медиана		Медиана	53,32
Мода	15000000			Мода		Мода	51,51
Стандартное отклонение		Стандартное отклонение		Стандартное отклонения		Стандартное отклонение	15,31459606
Дисперсия выборки		Дисперсия выборки		Дисперсия выборки		Дисперсия выборки	234,5368524
Эксцесс	14,30204373		0,348515174		10,62886069		21,84199947
Асимметричность		Асимметричность		Асимметричность		Асимметричность	3,92639839
Интервал		Интервал		Интервал		Интервал	132
Минимум		Минимум		Минимум		Минимум	30,1
Максимум	89724000	Максимум	2025	Максимум	4000	Максимум	162,1
Сумма	2195432200	Сумма	234976	Сумма	87150	Сумма	6578,86
Счет	117	Счет	117	Счет	117	Счет	117
Площадь кухни		Жилая площадь		Этаж		Этажей в доме	
т площаов кухна		лилал площаов		Omax		OTTAXEE & COME	
Среднее	15,31025641		25,88888889		14,21367521		24,66666667
Стандартная ошибка	0,567693438	Стандартная ошибка	0,690193592	Стандартная ошибка	0,978009691	Стандартная ошибка	1,139994467
Медиана	18,1	Медиана	21,9	Медиана	12	Медиана	26
Мода		Мода		Мода		Мода	34
Стандартное отклонение		Стандартное отклонение		Стандартное отклонения		Стандартное отклонение	12,33092552
Дисперсия выборки		Дисперсия выборки		Дисперсия выборки		Дисперсия выборки	152,0517241
Эксцесс	-0,055388398		3,398719542		-0,655001385		1.603637072
Асимметричность		Асимметричность		Асимметричность		Асимметричность	0.62038737
Интервал		Интервал		Интервал		Интервал	69
							5
Минимум		Минимум		Минимум		Минимум	
Максимум		Максимум		Максимум		Максимум	74
Сумма		Сумма		Сумма		Сумма	2886
Счет	117	Счет	117	Счет	117	Счет	117
Вид объекта		Балкон/Лоджия		Лифт		Парковка	
Среднее	0,564102564	1 Сполиоо	0,367521368	Сполиос	0,98290598	Сполиоо	0.641025641
Стандартная ошибка		В Стандартная ошибка		1 Стандартная ошибка		2 Стандартная ошибка	0,044538996
Медиана		1 Медиана) Медиана		1 Медиана	0,044530556
Мода		1 Мода		Мода		1 Мода	
Стандартное отклонение		7 Стандартное отклонение		Стандартное отклонения		В Стандартное отклонение	0,481762899
Дисперсия выборки		1 Дисперсия выборки		Дисперсия выборки		5 Дисперсия выборки	0,232095491
Эксцесс	-1,965295708		-1,71976118		55,929981		-1,674131415
Асимметричность		4 Асимметричность		Э Асимметричность		Э Асимметричность	-0,595638482
Интервал		1 Интервал		1 Интервал		1 Интервал	1
Минимум	() Минимум	() Минимум		0 Минимум	(
Максимум	•	1 Максимум		1 Максимум		1 Максимум	1
Сумма	66	6 Сумма	43	3 Сумма	11	5 Сумма	75
Счет		7 Счет		7 Счет		7 Счет	117
		AMAR					
Ремонт		МКАД		Первый/последний			
Среднее	0.000340030	Ореднее	0,931623932		0,05128205		
Стандартная ошибка							
		5 Стандартная ошибка		Стандартная ошибка	0,02047963		
Медиана	0,03961495		0,023433836	6 Стандартная ошибка 1 Медиана		7	
Медиана Мода	0,03961495	5 Стандартная ошибка	0,023433836		,		
Мода	0,039614956 (5 Стандартная ошибка) Медиана) Мода	0,023433836	1 Медиана 1 Мода		0	
Мода Стандартное отклонение	0,039614958 (0,42850128	5 Стандартная ошибка) Медиана) Мода 5 Стандартное отклонение	0,023433836	1 Медиана 1 Мода 2 Стандартное отклонениє	0,22152114)) 6	
Мода Стандартное отклонение Дисперсия выборки	0,039614958 (0,42850128 0,183613322	5 Стандартная ошибка) Медиана) Мода 5 Стандартное отклонение 2 Дисперсия выборки	0,023433836 0,253475692 0,064249926	1 Медиана 1 Мода 2 Стандартное отклонениє 5 Дисперсия выборки	0,22152114 0,04907161	0 0 6 3	
Мода Стандартное отклонение Дисперсия выборки Эксцесс	0,039614955 (0,42850125 0,183613322 -0,47607755	5 Стандартная ошибка D Медиана D Мода 5 Стандартное отклонение 2 Дисперсия выборки D Эксцесс	0,023433836 0,253475692 0,064249926 10,17907123	1 Медиана 1 Мода 2 Стандартное отклонениє 5 Дисперсия выборки 3 Эксцесс	0,22152114 0,04907161 15,2488094	0 0 6 8	
Мода Стандартное отклонение Дисперсия выборки Эксцесс Асимметричность	0,039614955 (0,42850125 0,183613322 -0,47607755 1,237883862	5 Стандартная ошибка) Медиана) Мода 5 Стандартное отклонение 2 Дисперсия выборки 3 Эксцесс 2 Асимметричность	0,023433836 0,253475692 0,064249926 10,17907123 -3,464872168	1 Медиана 1 Мода 2 Стандартное отклонениє 6 Дисперсия выборки 3 Эксцесс 3 Асимметричность	0,22152114 0,04907161 15,2488094 4,12169896	0 0 6 6 8 6	
Мода Стандартное отклонение Дисперсия выборки Эксцесс Асимметричность Интервал	0,039614958 0,042850128 0,183613322 -0,47607755 1,237883862	5 Стандартная ошибка) Медиана) Мода 5 Стандартное отклонение 2 Дисперсия выборки Э Эксцесс 2 Асимметричность I Интервал	0,023433836 0,253475692 0,064249926 10,17907123 -3,464872166	1 Медиана 1 Мода 2 Стандартное отклонениє 6 Дисперсия выборки 3 Эксцесс 3 Асимметричность 1 Интервал	0,22152114 0,04907161 15,2488094 4,12169896	0 0 5 8 8 5 9	
Мода Стандартное отклонение Дисперсия выборки Эксцесс Асимметричность Интервал Минимум	0,039614958 0,42850128 0,183613322 -0,47607758 1,237883862	5 Стандартная ошибка) Медиана) Мода 5 Стандартное отклонение 2 Дисперсия выборки 9 Эксцесс 2 Асимметричность 1 Интервал) Минимум	0,023433836 0,253475692 0,064249922 10,17907123 -3,464872166	І Медиана І Мода І Мода 2 Стандартное отклонениє 5 Дисперсия выборки 3 Эксцесс 3 Ассимметричность 1 Интервал 1 Минимум	0,22152114 0,04907161 15,2488094 4,12169896	0 0 6 8 5 9 1	
Мода Стандартное отклонение Дисперсия выборки Эксцесс Асимметричность Интервал Минимум Максимум	0,039614958 (0,042850124 0,183613322 -0,47607758 1,23788386	5 Стандартная ошибка) Медиана) Омода 5 Стандартное отклонение 2 Дисперсия выборки 3 Эксцесс 2 Асимметричность 1 Интервал 1 Минимум 1 Максимум	0,023433836 0,253475692 0,064249920 10,17907123 -3,464872160	Медиана Мода Стандартное отклонениє Дисперсия выборки Эксцесс Асимметричность Интервал Минимум Максимум	0,22152114 0,04907161: 15,2488094: 4,12169896	0 0 6 8 8 9 1	
Мода Стандартное отклонение Дисперсия выборки Эксцесс Асимметричность Интервал Минимум	0,039614958 (0,039614958) 0,42850128 0,183613322 -0,47607755 1,237883862	5 Стандартная ошибка) Медиана) Мода 5 Стандартное отклонение 2 Дисперсия выборки 9 Эксцесс 2 Асимметричность 1 Интервал) Минимум	0,023433836 0,253475692 0,064249920 10,17907123 -3,464872166	І Медиана І Мода І Мода 2 Стандартное отклонениє 5 Дисперсия выборки 3 Эксцесс 3 Ассимметричность 1 Интервал 1 Минимум	0,22152114 0,04907161: 15,2488094: 4,12169896	0 0 5 3 5 9 1 1	

Расстояние до метро

(Лист Описательная статистика в Данные.xlsx)

Из таблицы можно заметить, что:

- В выборке чуть больше новостроек (среднее 0,56)
- В большинстве квартир отсутствует балкон или лоджия (среднее 0,37)
- Почти во всех домах есть лифт (среднее 0,98)
- Практически все квартиры не находятся на первом или на последнем этаже (среднее 0,05)
- В большинстве жк есть парковка (среднее 0,64)
- Большинство не нуждается в ремонте (0,24)
- Практически все квартиры находятся в пределах МКАД (среднее 0,93)

Корреляция

	Цена	Вид объекта	Балкон/Лоджия	Пифп	Парковка	Ремонт	МКАД	Первый/последний	Год постройки	Расстояние до метро	Площадь	Площадь кухни	Жилая площадь	Этаж	Этажей в доме
Цена	1														
Вид объекта	-0,19	1													
Балкон/Лоджия	0,06	-0,72	1												
Лифт	0,05	0,15	-0,17	1											
Парковка	0,16	0,49	-0,46	0,18	1										
Ремонт	0,18	-0,27	0,24	-0,24	-0,29	1									
МКАД	0,14	0,04	-0,07	-0,04	0,15	-0,09	1								
Первый/последний	-0,07	-0,26	0,14	-0,27	-0,15	0,41	-0,09	1							
Год постройки	0,06	0,76	-0,60	0,23	0,56	-0,21	-0,07	-0,20	1						
Расстояние до метро	-0,10	-0,24	0,24	0,06	-0,45	0,14	-0,52	0,16	-0,16	1					
Площадь	0,80	-0,04	0,06	0,11	0,21	0,16	0,07	-0,04	0,13	-0,04	1				
Площадь кухни	0,29	0,60	-0,51	0,21	0,54	-0,15	0,11	-0,25	0,68	-0,24	0,45	1			
Жилая площадь	0,54	-0,45	0,29	0,06	-0,13	0,33	-0,07	0,23	-0,24	0,18	0,52	-0,18	1		
Этаж	0,07	0,51	-0,49	0,12	0,34	-0,24	0,10	-0,20	0,55	-0,30	0,01	0,37	-0,15	5 1	
Этажей в доме	0.13	0,66	-0.66	0.21	0.46	-0.27	0.17	-0.28	0,71	-0,35	0.09	0.47	-0.14	0.76	1

(Лист Корреляция в Данные.xlsx)

Из таблицы можно заметить высокую корреляцию (|x| > 0.7) между следующими параметрами:

- Год постройки и Вид объекта (год постройки новостроек больше)
- Год постройки и Этажей в доме (чем позже построен дом, тем больше в нем этажей)
- Балкон/Лоджия и Вид объекта (в новостройках балкон встречается реже)
- Цена и Площадь (чем больше площадь, тем выше цена)
- Этаж и этажей в доме (чем больше общее число этажей, тем большим может быть этаж, на котором находится квартира)

Линейная регрессионная модель

Первая

ВЫВОД ИТОГОВ						
Регрессионная ст	атистика					
Множественный R	0,82					
R-квадрат	0,67					
Нормированный R-квадрат	0,66					
Стандартная ошибка	7565687,41					
Наблюдения	117					
Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	5	1,31E+16	2,62E+15	4,58E+01	1,93E-25	
Остаток	111	6,35E+15	5,72E+13			
Итого	116	1,95E+16				
	Коэффициенты	Стандартная ошибка	t-cmamucmuкa	Р-Значение	Нижние 95%	Верхние 95%
Ү-пересечение	-47289586,15	83509482,79	-0,57	0,57	-212769201,14	118190028,83
Год постройки	13446,14	42131,86	0,32	0,75	-70040,95	96933,23
Расстояние до метро	-2253,68	1155,59	-1,95	0,05	-4543,57	36,21
Площадь	613852,64	72596,27	8,46	0,00	469998,28	757707,00
Площадь кухни	-108915,97	200711,99	-0,54	0,59	-506640,19	288808,24
Жилая площадь	304342,21	133171,99	2,29	0,02	40453,01	568231,41

(Лист Линейная 1 в Данные.xlsx)

 $R^2 = 0.67$, что говорит о неплохой описательной способности модели.

По оценке можно заметить, что p-value регрессора Площадь равен 0 => на любом уровне значимости коэффициент значим.

P-value Расстояния до метро и Жилой площади также достаточно низки => на уровне значимости выше 5% коэффициенты значимы.

P-value Площади кухни и Года постройки достаточно высоки из-за корреляции. В итоговой модели для решения проблемы мультиколлинеарности стоит их исключить.

Интерпретация коэффициентов:

При увеличении года постройки на 1 год цена на квартиру увеличивается на 13446,14 Р.

При увеличении расстояния до метро на 1 метр цена на квартиру уменьшается на 2253,68 Р.

При увеличении общей площади на 1 м^2 цена на квартиру увеличивается на 613852,64 P.

При увеличении площади кухни на 1 м 2 цена на квартиру уменьшается на 108915,97 m P.

При увеличении жилой площади на 1 м^2 цена на квартиру увеличивается на 304342,21 P.

Вторая

вывод итогов						
Регрессионная ст	атистика					
Множественный R	0,85					
R-квадрат	0,72					
Нормированный R-квадрат	0,70					
Стандартная ошибка	7147289,15					
Наблюдения	117					
Дисперсионный анализ						
	df	SS	MS	F	Значимость <i>F</i>	
Регрессия	10	1,40E+16	1,40E+15	2,75E+01	4,05E-25	
Остаток	106	5,41E+15	5,11E+13			
Итого	116	1,95E+16	-			
	Коэффициенты	Стандартная ошибка	t-статистика	Р-Значение	Нижние 95%	Верхние 95%
Ү-пересечение	-11131660,76			0,12	-25228480,28	2965158,77
Вид объекта	-8076627,47	2221268,84	-3,64	0,00	-12480508,99	-3672745,95
Балкон/Лоджия	-5568250,18	2059640,10		0,01	-9651687,01	-1484813,36
Лифт	-5263926,80	5594177,62	-0,94	0,35	-16354928,03	5827074,44
Парковка	1481211,80	1888193,66	0,78	0,43	-2262315,79	5224739,39
Ремонт	1964886,33	1864079,69	1,05	0,29	-1730832,96	5660605,62
МКАД	2344735,83	3125431,66	0,75	0,45	-3851736,42	8541208,08
Первый/последний	-8324610,16	3517884,57	-2,37	0,02	-15299158,48	-1350061,84
Жилая площадь	215467,91	126382,48		0,09	-35097,65	466033,46
Расстояние до метро	-1105,07			0,43	-3889,39	1679,24
Площадь	600043,01	56135,51	10,69	0,00	488748,90	711337,12

(Лист Линейная 2 в Данные.xlsx)

 $R^2 = 0.72$

По оценке можно заметить, что p-value регрессоров Площадь и Вид объекта равен 0 = > на любом уровне значимости коэффициенты значимы.

P-value Первого/последнего, Балкона/Лоджии и Жилой площади также достаточно низки => на уровне значимости выше 9% коэффициенты значимы.

P-value Лифта, Парковки, Ремонта, МКАД достаточно высоки из-за корреляции. В итоговой модели для решения проблемы мультиколлинеарности стоит их исключить.

Интерпретация коэффициентов:

Цена на квартиру в новостройке ниже на 8076627,47 Р.

Цена на квартиру, в которой есть балкон/лоджия, ниже на 5568250,18 ₽.

Цена на квартиру, в которой требуется ремонт, выше на 1964886,33 Р.

Цена на квартиру, находящуюся в здании в пределах МКАД, выше на 2344735,83 Р.

Цена на квартиру, находящуюся на первом или на последнем этаже здания, ниже на 8324610,16 P.

Цена на квартиру, находящуюся в ЖК, в котором есть парковка, выше на 1481211,8 ₽.

При увеличении расстояния до метро на 1 метр цена на квартиру уменьшается на 1105,07 Р.

При увеличении общей площади на 1 м^2 цена на квартиру увеличивается на 600043,01 P.

При увеличении жилой площади на 1 m^2 цена на квартиру увеличивается на 215467,91 P.

Третья

ВЫВОД ИТОГОВ						
Регрессионная стап	пистика					
Множественный R	0,85					
R-квадрат	0,71					
Нормированный R-квадрат	0,70					
Стандартная ошибка	7111310,49					
Наблюдения	117,00					
Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	6	1,38895E+16	2,31491E+15	45,775721	1,03369E-27	
Остаток	110	5,56278E+15	5,05707E+13			
Итого	116	1,94523E+16				
	Коэффициентьпа	ндартная ошиб	t-статистика	Р-Значение	Нижние 95%	Верхние 95%
Ү-пересечение	-13037024,48	3680352,62	-3,54	0,00	-20330619,53	-5743429,42
Вид объекта	-7977737,93	2144229,32	-3,72	0,00	-12227097,15	-3728378,71
Балкон/Лоджия	-5471611,98	2003429,16	-2,73	0,01	-9441938,33	-1501285,63
Первый/последний	-6099755,36	3167119,92	-1,93	0,06	-12376243,65	176732,94
Жилая площадь	211304,72	123335,17	1,71	0,09	-33116,63	455726,07
Расстояние до метро	-2077,83	1096,26	-1,90	0,06	-4250,35	94,69
Площадь	617160,61	53396,61	11,56	0,00	511341,07	722980,15

(Лист Линейная 3 в Данные.xlsx)

 $R^2 = 0.71$

По оценке можно заметить, что p-value всех регрессоров достаточно низки => на уровне значимости, больших 9%, все коэффициенты значимы.

Интерпретация коэффициентов:

Цена на квартиру в новостройке ниже на 7977737,93 ₽.

Цена на квартиру, в которой есть балкон/лоджия, ниже на 5471611,98 ₽.

Цена на квартиру, находящуюся на первом или на последнем этаже здания, ниже на 6099755,36 Р.

При увеличении расстояния до метро на 1 метр цена на квартиру уменьшается на 2077,83 Р.

При увеличении общей площади на 1 м 2 цена на квартиру увеличивается на 617160,61 \mathbb{P} .

При увеличении жилой площади на 1 м 2 цена на квартиру увеличивается на $211304,72 \ \mathbb{P}$.

Смешанная регрессионная модель

вывод итогов						
Регрессионная ста	тистика					
Множественный R	0,75					
R-квадрат	0,56					
Нормированный R-квадрат	0,54					
Стандартная ошибка	0,31					
Наблюдения	117					
Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	6	13,51	2,25	23,69	0,00	
Остаток	110	10,46	0,10			
Итого	116	23,97				
	Коэффициентыпан	дартная ошиб	t-статистика	Р-Значение	Нижние 95%	Верхние 95%
Ү-пересечение	10,71	0,62	17,28	0,00	9,48	11,94
Жилая площадь	0,22	0,15	1,46	0,15	-0,08	0,51
Расстояние до метро	-0,06	0,04	-1,32	0,19	-0,15	0,03
Площадь	1,43	0,17	8,59	0,00	1,10	1,76
Вид объекта	-0,18	0,09	-1,95	0,05	-0,36	0,00
Балкон/Лоджия	-0,10	0,09	-1,18	0,24	-0,27	0,07
Первый/последний	-0.23	0.14	-1,64	0,10	-0.50	0,05

(Лист Логарифмическая в Данные.xlsx)

 $R^2 = 0.56$

(Логарифмическая для регрессоров жилая площадь, расстояние до метро, площадь; Логарифмически-линейная для остальных)

По оценке можно заметить, что p-value у некоторых регрессоров немного увеличился. Теперь все коэффициенты значимы при уровне значимости выше 24%.

Интерпретация коэффициентов:

Цена на квартиру в новостройке ниже на 18%.

Цена на квартиру, в которой есть балкон/лоджия, ниже на 10%.

Цена на квартиру, находящуюся на первом или на последнем этаже здания, ниже на 23%.

При увеличении расстояния до метро на 1% цена на квартиру уменьшается на 0.06%.

При увеличении общей площади на 1% цена на квартиру увеличивается на 1,43%.

При увеличении жилой площади на 1% цена на квартиру увеличивается на 0,22%.

Тест Чоу

Эталонная модель:

Цена
$$_i = \alpha + \beta_1 *$$
 ВидОбъекта $_i + \beta_2 *$ БалконЛоджия $_i + \beta_3 *$ ПервыйПоследний $_i + \beta_4 *$ РасстояниеДоМетро $_i + \beta_5 *$ Площадь $_i + \beta_6 *$ ЖилаяПлощадь $_i$

В этой модели есть 3 dummy переменных, проверим модель на наличие структурных сдвигов.

1 тест (Вид объекта)

$$F_{\text{набл}} = 3,16; F_{\text{кр}} = 1,83$$

 $F_{\rm kp} < F_{\rm набл} => {\rm H_0}$ отвергается на уровне значимости 10%, цены на новостройки отличаются от цен на вторичное жилье, в модели есть структурный сдвиг.

2 тест (Балкон/Лоджия)

$$F_{\text{набл}} = 18,52; F_{\text{кр}} = 1,83$$

 $F_{\rm kp} < F_{\rm Haбл} => H_0$ отвергается на уровне значимости 10%, цены на квартиры, в которых есть балкон/лоджия, отличаются от цен на квартиры без них, в модели есть структурный сдвиг.

3 тест (Первый/Последний)

$$F_{\text{набл}} = 0,998; F_{\text{кр}} = 1,83$$

 $F_{\rm кр} > F_{\rm набл} = > {\rm H}_0$ не отвергается на уровне значимости 10%, цены на квартиры на последнем или на первом этаже не отличаются от цен на другие квартиры, в модели нет структурного сдвига.

Тест Бокса-Кокса

(1) Цена $_i = \alpha + \beta_1 *$ ВидОбъекта $_i + \beta_2 *$ БалконЛоджия $_i + \beta_3 *$ ПервыйПоследний $_i + \beta_4 *$ РасстояниеДоМетро $_i + \beta_5 *$ Площадь $_i + \beta_6 *$ ЖилаяПлощадь $_i$

(2) $\ln(\text{Цена}_i) = \alpha + \beta_1 * \text{ВидОбъекта}_i + \beta_2 * \text{БалконЛоджия}_i + \beta_3 *$ ПервыйПоследний $_i + \beta_4 * \text{РасстояниеДоМетро}_i + \beta_5 * Площадь<math>_i + \beta_6 *$ ЖилаяПлощадь $_i$

 H_0 : модели (1) и (2) имеют одинаковое качество

 H_1 : модели (1) и (2) имеют разные качество, то есть одна из моделей лучше другой

Цена
$$_i^* = \frac{$$
 Цена $_i$ $}{\sqrt[n]{$ Цена $_1*...*$ Цена $_n$

Цена $_i^* = \alpha + \beta_1 *$ ВидОбъекта $_i + \beta_2 *$ БалконЛоджия $_i + \beta_3 *$ ПервыйПоследний $_i + \beta_4 *$ РасстояниеДоМетро $_i + \beta_5 *$ Площадь $_i + \beta_6 *$ ЖилаяПлощадь $_i$

Оценим модель: $RSS_1^* = 22,96$.

$$\begin{split} \ln(\text{Ценa}_i^*) &= \alpha + \beta_1 * \text{ВидОбъекта}_i + \beta_2 * \text{БалконЛоджия}_i + \beta_3 \\ &* \text{ПервыйПоследний}_i + \beta_4 * \text{РасстояниеДоМетро}_i + \beta_5 \\ &* \text{Площадь}_i + \beta_6 * \text{ЖилаяПлощадь}_i \end{split}$$

Оценим модель: $RSS_2^* = 10,57$.

$$T_{\text{набл}} = \frac{n}{2} * \left| \ln \left(\frac{RSS_2^*}{RSS_1^*} \right) \right| = 68,7$$

Рассчитаем критическую точку на уровне значимости 5%: $T_{\text{кp}} = chi2inv(0,95;1) = 3,8415$.

ВЫВОД: $T_{\text{набл}} > T_{\text{кр}} = > модели$ (1) и (2) имеют разное качество.

Приложение

Расчеты для теста Чоу

