

Министерство науки и высшего образования
Российской Федерации

Федеральное Государственное
Автономное Образовательное Учреждение
Высшего Образования
Национальный ядерный университет «МИФИ»

Кафедра: «Финансовый мониторинг»

Отчет по Лабораторной работе №4:

Студент Монастырский М. О.

Группа С21-703

Проверила: Домашова Д. В.

Оглавление

Введение	3
Постановка задачи.....	4
Порядок выполнения	5
Критерий Чоу.....	8
Введение фиктивных переменных	10
Построение модели с фиктивными переменными.....	11
Выводы	12

Введение

О линейных регрессионных моделях с переменной структурой будем говорить в тех случаях, когда на результативную переменную помимо отобранных и измеренных объясняющих признаков оказывает существенное воздействие меняющиеся (одновременно с предопределенными переменными во времени и/или в пространстве) некоторые качественные факторы, что может вести к скачкообразным изменениям коэффициентов линейной регрессии.

Очевидна идея, связанная с разбиением исходных статистических данных на качественно-однородные группы и последующей оценкой функции регрессии в каждой из таких групп. Но такой подход либо ведет к снижению статистической надежности результатов, либо невозможен, ввиду малого объема выборки, хотя бы в одной из регрессионно-однородных подвыборок.

Выход заключается в введение фиктивных переменных (“манекенов”), однако, следует обоснованно подходить к их введению, поскольку каждая новая переменная ведет к уменьшению степеней свободы и снижению надежности выводов. Приобретение навыков построения и анализа эконометрических моделей по регрессионно-неоднородным данным является целью предлагаемой работы.

Постановка задачи

По данным:

Наименование показателя	Обозначение
Административный округ Москвы	X1
Площадь жилой (кв.м)	X2
Общая площадь (кв.м)	X4
Число комнат (/2/3)	X5
Количество этажей в доме	X9
Этаж квартиры (первый/промежуточный/последний)	X10
Тип дома (кирпичный/монолитный/панельный/блочный/кирпично-монолитный/сталинский)	X12
Вид из окна (во двор/на улицу/во двор и на улицу)	X19
Цена квартиры (тыс. руб.)	Y

1. Выдвинуть и обосновать предположение о сопутствующих качественных факторах, числе уровней каждого, указать число фиктивных переменных и охарактеризовать каждую из них.
2. Записать линейную модель регрессии с переменной структурой и её матрицу “объект - свойства”.
3. Исследовать имеющиеся статистические данные на неоднородность с помощью критерия Чоу.
4. Оценить параметры регрессионной модели с переменной структурой и провести её анализ.

Порядок выполнения

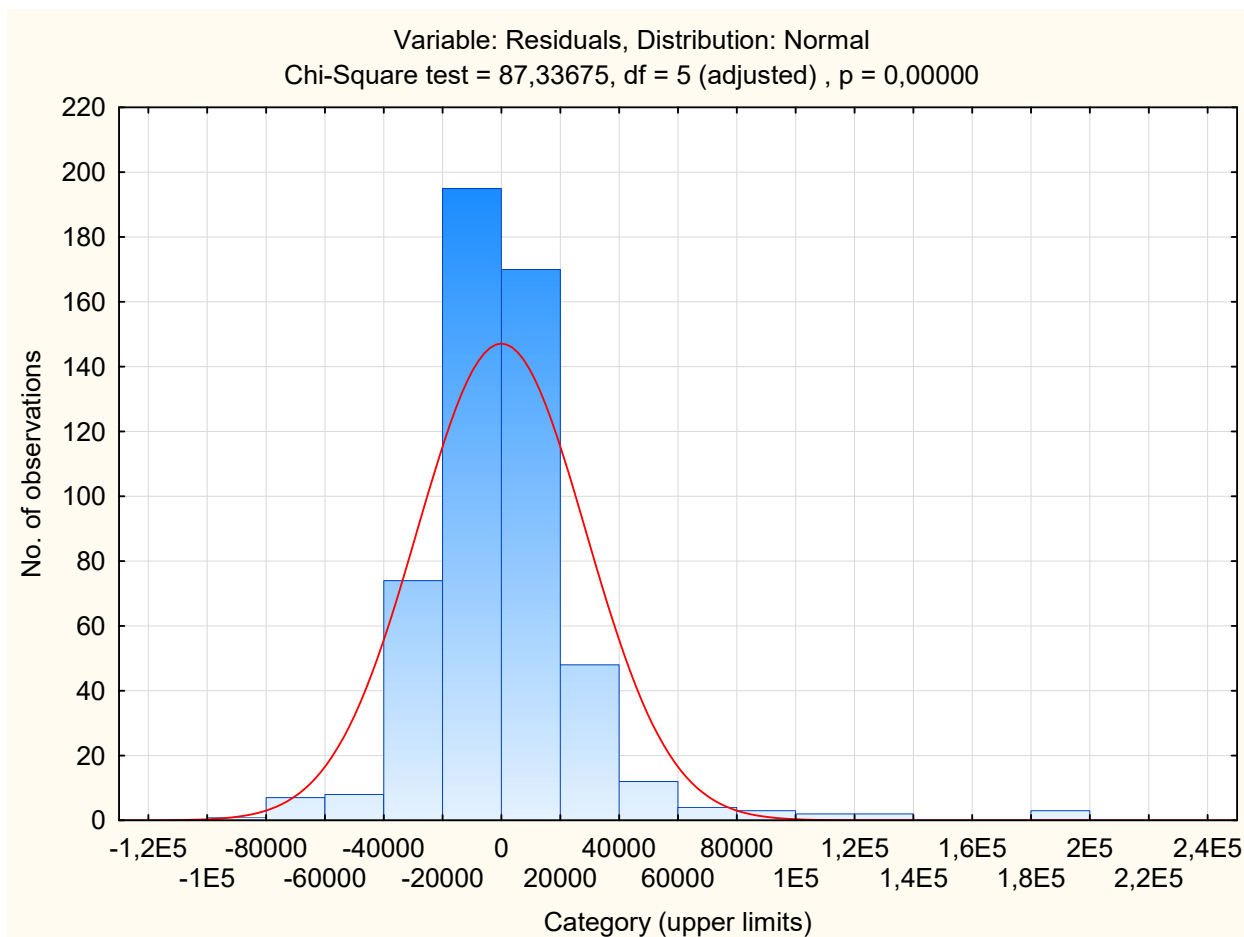
Построим линейную регрессионную модель для нашей выборки, используя в качестве объясняющих признаков только количественные переменные (X_2 , X_4 , X_9):

Multiple Regression Results

Dependent: Y Multiple R = ,79188376 F = 294,2694
 R² = ,62707990 df = 3,525
 No. of cases: 529 adjusted R² = ,62494892 p = 0,000000
 Standard error of estimate: 28769,357570
 Intercept: -42787,39323 Std.Error: 3276,318 t(525) = -13,06 p = 0,0000
 X2 b* = ,226 X4 b* = ,588 X9 b* = -,05

(significant b* are highlighted in red)

N=52 9	Regression Summary for Dependent Variable: Y (Sheet1 in Лаб 4 данные.stw) R= ,79188376 R²= ,62707990 Adjusted R²= ,62494892 F(3,525)=294,27 p					
	b*	Std.Err. of b*	b	Std.Err. of b	t(525)	p-value
Intercept			-42787,4	3276,318	-13,0596	0,000000
X2	0,226314	0,058995	488,3	127,300	3,8362	0,000140
X4	0,587822	0,058802	817,8	81,809	9,9966	0,000000
X9	-0,046506	0,026915	-159,6	92,380	-1,7279	0,084593



$P = 0.00 < 0.05$

Модель значима

Поскольку можно предположить нормальный характер распределения регрессионных остатков, то на основании отчета делаем выводы:

- модель значима;

- значимое влияние на результативный признак – цена квартиры, оказывают объясняющие переменные – общая площадь квартиры (X4), жилая площадь квартиры (X2)

- оценка уравнения регрессии:

$$\hat{y} = -42787,4_{3276,318} + 488,3_{12,3}x_2 + 817,8_{81,809}x_4$$

На результативный признак могут также влиять качественные переменные $x_1, x_{10}, x_{12}, x_{19}$. Прежде чем включить их в модель, проверим выборочную совокупность на однородность с помощью критерия Чоу.

Проверим сначала по переменной x_1 административный округ Москвы.

$H_0: \begin{matrix} \beta^{(1)} = \beta^{(2)} \\ \sigma_{\varepsilon(1)}^2 = \sigma_{\varepsilon(2)}^2 \end{matrix}$ Проверяется с помощью статистики:

$$\gamma_{n_1 n_2} = \frac{(e^T \cdot e - e^{(1)T} \cdot e^{(1)} - e^{(2)T} \cdot e^{(2)}) / (k + 1)}{\frac{e^{(1)T} \cdot e^{(1)} + e^{(2)T} \cdot e^{(2)}}{n_1 + n_2 - 2k - 2}}$$

Критерий Чоу

В условиях справедливости нулевой гипотезы эта статистика распределена по закону Фишера – Снедекора с $\nu_1 = k + 1$ и $\nu_2 = n_1 + n_2 - 2k - 2$

- 1) Разбиваем выборку на 6 однородных группы. Построим уравнение по объединенной выборки и получим:

Analysis of Variance; DV: Y (Sheet1 in Лаб 4 данные.stw)					
Effect	Sums of Squares	df	Mean Squares	F	p-value
Regress.	7,306791E+11	3	2,435597E+11	294,2694	0,00
Residual	4,345299E+11	525	8,276759E+08		
Total	1,165209E+12				

$$Q_0 = \sum_{i=1}^n e_i^2 = e^T e = 4,345299E + 11$$

Analysis of Variance; DV: Y (Sheet1 in Лаб 4 данные.stw) Include condition: X1 = "южный"					
Effect	Sums of Squares	df	Mean Squares	F	p-value
Regress.	4,977996E+09	3	1,659332E+09	67,31178	0,000000
Residual	4,166093E+09	169	2,465144E+07		
Total	9,144089E+09				

$$Q_1 = \sum_{i=1}^n e_i^2 = e^T e = 4,166093E + 09$$

Analysis of Variance; DV: Y (Sheet1 in Лаб 4 данные.stw) Exclude condition: X1 = "южный"					
Effect	Sums of Squares	df	Mean Squares	F	p-value
Regress.	6,573540E+11	3	2,191180E+11	199,2668	0,00
Residual	3,870666E+11	352	1,099621E+09		
Total	1,044421E+12				

$$Q_2 = \sum_{i=1}^n e_i^2 = e^T e = 3,870666E + 11$$

$$\gamma_{n_1 n_2} = \frac{(e^T \cdot e - e^{(1)T} \cdot e^{(1)} - e^{(2)T} \cdot e^{(2)}) / (k+1)}{(e^{(1)T} \cdot e^{(1)} + e^{(2)T} \cdot e^{(2)}) / (n_1 + n_2 - 2k - 2)} = \frac{(4,345299E+11 - 4,166093E+09 - 3,870666E+11) / (3+1)}{(4,166093E+09 + 3,870666E+11) / (529 - 2 \cdot 3 - 2)} = 5,31E-05$$

$$\text{Расч} = F(4;521) = 2,3289$$

Выборки неоднородны

Effect	Analysis of Variance; DV: Y (Sheet1 in Лаб 4 данные.stw) Include condition: X1 = "северо-восточный"				
	Sums of Squares	df	Mean Squares	F	p-value
Regress.	8,756630E+09	3	2,918877E+09	68,50909	0,000000
Residual	7,285572E+09	171	4,260568E+07		
Total	1,604220E+10				

Effect	Analysis of Variance; DV: Y (Sheet1 in Лаб 4 данные.stw) Exclude condition: X1 = "северо-восточный"				
	Sums of Squares	df	Mean Squares	F	p-value
Regress.	6,666043E+11	3	2,222014E+11	203,4358	0,00
Residual	3,822853E+11	350	1,092244E+09		
Total	1,048890E+12				

5,54E-05 < 2,389045 выборки неоднородны

Effect	Analysis of Variance; DV: Y (Sheet1 in Лаб 4 данные.stw) Include condition: X1 = "центральный"				
	Sums of Squares	df	Mean Squares	F	p-value
Regress.	4,155120E+11	3	1,385040E+11	79,32048	0,000000
Residual	3,090653E+11	177	1,746132E+09		
Total	7,245773E+11				

0,00017 < 2,389045 Выборки неоднородны

Effect	Analysis of Variance; DV: Y (Sheet1 in Лаб 4 данные.stw) Exclude condition: X1 = "центральный"				
	Sums of Squares	df	Mean Squares	F	p-value
Regress.	1,335026E+10	3	4,450085E+09	127,8497	0,00
Residual	1,197367E+10	344	3,480717E+07		
Total	2,532392E+10				

2) Разбиваем выборку на 2 однородных группы для оценки критерия ЧОУ по переменной X5

Effect	Analysis of Variance; DV: Y (Sheet1 in Лаб 4 данные.stw) Include condition: X5=2				
	Sums of Squares	df	Mean Squares	F	p-value
Regress.	1,109541E+11	3	3,698469E+10	70,52105	0,000000
Residual	1,311123E+11	250	5,244490E+08		
Total	2,420663E+11				

Effect	Analysis of Variance; DV: Y (Sheet1 in Лаб 4 данные.stw) Exclude condition: X5=2				
	Sums of Squares	df	Mean Squares	F	p-value

Regress.	6,046567E+11	3	2,015522E+11	225,7720	0,00
Residual	2,419284E+11	271	8,927245E+08		
Total	8,465851E+11				

7,91E-05 < 2,389045 Значения неоднородны

Введение фиктивных переменных

Наименование показателя	Обозначение
Административный округ Москвы	X1
Площадь жилой (кв.м)	X2
Общая площадь (кв.м)	X4
Число комнат (2/3)	X5
Количество этажей в доме	X9
Этаж квартиры (первый/промежуточный/последний)	X10
Тип дома (кирпичный/монолитный/панельный/блочный/кирпично-монолитный/сталинский)	X12
Вид из окна (во двор/на улицу/во двор и на улицу)	X19
Цена квартиры (тыс. руб.)	Y

X₁ имеет 3 градации, вводим 2 фиктивные переменные (Центральный-базовая):

$$d_1^{(1)} = \begin{cases} 1, \text{ Южный округ} \\ 0, \text{ иначе} \end{cases}$$

$$d_2^{(1)} = \begin{cases} 1, \text{ Северо – Восточный} \\ 0, \text{ иначе} \end{cases}$$

X₁₀ имеет 3 градации, вводим 2 фиктивные переменные (первый-базовая):

$$d_1^{(10)} = \begin{cases} 1, \text{ промежуточный} \\ 0, \text{ иначе} \end{cases}$$

$$d_2^{(10)} = \begin{cases} 1, \text{ последний} \\ 0, \text{ иначе} \end{cases}$$

X₅ имеет 2 градации, вводим 1 фиктивные переменные (первый-двухкомнатная):

$$d_1^{(5)} = \begin{cases} 1, \text{ если в наблюдении } i \text{ квартира трехкомнатная} \\ 0 \text{ иначе} \end{cases}$$

X₁₉ имеет 3 градации, вводим 2 фиктивные переменные (во двор-базовая):

$$d_1^{(19)} = \begin{cases} 1, \text{ на улицу} \\ 0, \text{ иначе} \end{cases}$$

$$d_2^{(19)} = \begin{cases} 1, \text{ во двор и на улицу} \\ 0, \text{ иначе} \end{cases}$$

X₁₂ имеет 6 градаций, вводим 5 фиктивных переменных (кирпичный-базовая):

$$d_1^{(12)} = \begin{cases} 1, \text{ монолитный} \\ 0, \text{ иначе} \end{cases}$$

$$d_2^{(12)} = \begin{cases} 1, \text{ панельный} \\ 0, \text{ иначе} \end{cases}$$

$$d_3^{(12)} = \begin{cases} 1, \text{ блочный} \\ 0, \text{ иначе} \end{cases}$$

$$d_4^{(12)} = \begin{cases} 1, \text{ кирпично – монолитный} \\ 0, \text{ иначе} \end{cases}$$

$$d_5^{(12)} = \begin{cases} 1, \text{ сталинский} \\ 0, \text{ иначе} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_{i2} + \beta_2 x_{i4} + \beta_3 x_{i9} + \beta_4 d_{i1}^{(1)} + \beta_5 d_{i2}^{(1)} + \beta_6 d_{i1}^{(5)} + \beta_7 d_{i1}^{(10)} + \beta_8 d_{i2}^{(10)} + \beta_9 d_{i1}^{(12)} + \beta_{10} d_{i2}^{(12)} + \beta_{11} d_{i3}^{(12)} + \beta_{12} d_{i4}^{(12)} + \beta_{13} d_{i5}^{(12)} + \beta_{14} d_{i2}^{(19)} + \beta_{15} d_{i1}^{(19)}$$

Построение модели с фиктивными переменными

Regression Summary for Dependent Variable: Y (Sheet1 in фиктивные) R= ,84354164 R?= ,71156249 Adjusted R?= ,70312865 F(15,513)=84,370 p						
N=529	b*	Std.Err. of b*	b	Std.Err. of b	t(513)	p-value
Interc ept			-10506,1	6406,08	-1,64001	0,101615
X1_1	-0,277318	0,034927	-27743,4	3494,17	-7,93992	0,000000
X1_2	-0,277262	0,034343	-27656,6	3425,73	-8,07320	0,000000
X2	0,268090	0,056822	578,5	122,61	4,71809	0,000003
X4	0,457929	0,058730	637,1	81,71	7,79717	0,000000
X5_1	-0,119610	0,030628	-11236,1	2877,18	-3,90525	0,000107
X10_ 1	-0,052112	0,042815	-6067,5	4985,07	-1,21714	0,224109
x10_2	-0,039623	0,042687	-5244,9	5650,57	-0,92821	0,353733
X12_ 1	0,153965	0,036851	15449,4	3697,78	4,17801	0,000035
X12_ 2	0,090861	0,035824	9189,6	3623,21	2,53632	0,011498
X12_ 3	0,032137	0,027574	6150,2	5276,97	1,16547	0,244368
X12_ 4	0,066035	0,029384	10998,9	4894,21	2,24733	0,025043
X12_ 5	-0,008792	0,024021	-6723,4	18369,46	-0,36601	0,714509
X19_ 2	0,023580	0,026959	2338,9	2674,03	0,87469	0,382152
X19_ 1	0,027829	0,027554	3147,0	3115,83	1,01000	0,312973

X9	-0,118049	0,027638	-405,2	94,86	-4,27127	0,000023
----	-----------	----------	--------	-------	----------	----------

После исключения мультиколлинеарности методом включения переменных получаем

Regression Summary for Dependent Variable: Y (Sheet1 in фиктивные) R= ,84209304 R?= ,70912069 Adjusted R?= ,70407654 F(9,519)=140,58 p						
N=529	b*	Std.Err. of b*	b	Std.Err. of b	t(519)	p-value
Intercept			-13716,3	4494,937	-3,05150	0,002394
X4	0,455994	0,057993	634,4	80,683	7,86284	0,000000
X5_1	-0,119291	0,030310	-11206,1	2847,294	-3,93570	0,000094
X2	0,272093	0,056283	587,1	121,448	4,83438	0,000002
X1_2	-0,269451	0,033221	-26877,5	3313,762	-8,11087	0,000000
X1_1	-0,268015	0,033273	-26812,7	3328,657	-8,05512	0,000000
X9	-0,114119	0,026324	-391,7	90,351	-4,33523	0,000017
X12_1	0,134203	0,033979	13466,4	3409,548	3,94961	0,000089
X12_2	0,068408	0,031714	6918,8	3207,552	2,15704	0,031461
X12_4	0,055374	0,027979	9223,2	4660,246	1,97912	0,048330

$$\hat{y} = -13716,3 + 634,4 x_4 + 587,1 x_2 - 11206,1 d_1^{(5)} - 26812,7 d_1^{(1)} - 26877,5 d_2^{(1)} - 391,7 x_9 + 13466,4 d_1^{(12)} + 6918,8 d_2^{(12)} + 9223,2 d_4^{(12)}$$

Выводы

В ходе работы на основе данных, была проверена однородность выборки по всем качественным переменным и были введены фиктивные переменные, перечисленные в пункте 4 работы, а также построена следующая регрессионная модель:

$$\hat{y} = -13716,3 + 634,4 x_4 + 587,1 x_2 - 11206,1 d_1^{(5)} - 26812,7 d_1^{(1)} - 26877,5 d_2^{(1)} - 391,7 x_9 + 13466,4 d_1^{(12)} + 6918,8 d_2^{(12)} + 9223,2 d_4^{(12)}$$

Модель значима. После включения в модель фиктивной переменной значение коэффициента детерминации увеличилось с 0,62707990 до 0,709121, следовательно, учёт качественных признаков улучшили модель.

Полученная модель позволяет сделать следующие выводы:

- При увеличении жилой площади в квартире 1 кв.м. стоимость квартиры увеличивается на 634,4 тыс. руб.;
- При увеличении жилой площади на 1 кв.м. стоимость квартиры увеличивается на 587,1 тыс. руб.;

- С ростом этажности дома стоимость квартиры падает на 391,7 тыс рублей/ этаж
- Трехкомнатные квартиры стоят дешевле чем двухкомнатные на 11236,1 тыс рублей
- Квартиры в Южном округе стоят в среднем на 26812,7тыс. руб. дешевле, чем квартиры в Центральном;
- Квартиры в СВАО стоят в среднем 26877,5 тыс. руб. дешевле, чем квартиры в Центральном
- Квартиры в монолитном доме в среднем стоят на 13466,4 тыс рублей дороже чем в кирпичном
- Квартиры в панельном доме стоят в среднем на 6918,8 тыс рублей дороже чем в кирпичном
- Квартиры в кирпично-монолитном доме стоят в среднем на 9223,2 дороже чем в кирпичном