ЛАБОРАТОРНАЯ РАБОТА №4
ИССЛЕДОВАНИЕ СТРУКТУРНОЙ УСТОЙЧИВОСТИ КОЭФФИЦИЕНТОВ РЕГРЕССИИ С ПОМОЩЬЮ ТЕСТА ЧОУ

Цель: освоить методику проверки гипотезы о совпадении уравнений регрессии для двух выборок.

Задачи:

1. Построить модель линейной регрессии для двух выборок;
2. применить тест Чоу для проверки гипотезы о совпадении уравнений регрессии для двух выборок.

Основные сведения

Предположим, у нас есть две выборки, объема n_1 и n_2 соответственно для одного и того же набора переменных. По каждой выборке мы оцениваем коэффициенты уравнения регрессии:

$$Y_i = a_1 + a_2 X_{2i} + \ldots + a_k X_{ki} + \varepsilon_i', i = 1, \ldots, n_1,$$

$$Y_i = b_1 + b_2 X_{2i} + \ldots + b_k X_{ki} + \varepsilon_i'', i = n_1 + 1, \ldots, n_1 + n_2.$$

Нас интересует ответ на вопрос: что лучше, оценивать модель по объединенной выборке или по каждой выборке в отдельности?

Для проверки гипотезы $H_0: a_i = b_i, i = 1,2, \ldots, k, \sigma_{\varepsilon'}^2 = \sigma_{\varepsilon''}^2$ (т. е. лучше оценивать модель по объединенной выборке) при альтернативной (т. е. лучше оценивать модели по каждой выборке в отдельности) используется тест Чоу.

Тестовая статистика имеет вид:

$$F = \frac{(RSS - RSS_1 - RSS_2)/k}{(RSS_1 + RSS_2)/(n_1 + n_2 - 2k)},$$

где RSS – сумма квадратов остатков регрессии, оцененной по всем наблюдениям, RSS_1 – сумма квадратов остатков, оцененных по n_1 наблюдениям, RSS_2 – сумма квадратов остатков, оцененных по n_2 наблюдениям.

При выполнении нулевой гипотезы тестовая статистика имеет F – распределение со степенями свободы $(k, n_1 + n_2 - 2k)$. Если рассчитанное значение F – статистики не превышает критическое, то основная гипотеза не отвергается, зависимость можно считать единой для двух наборов данных.

Порядок выполнения лабораторной работы

Для выполнения данной работы студенту необходимо выбрать свой вариант из таблицы 3. Номер варианта определяется согласно номеру студента по списку (уточняется у преподавателя). Рассмотрим пример. Имеются две выборки объемами $n_1 = 11$ и $n_2 = 16$ соответственно (таблицы 1 и 2). Необходимо проверить гипотезу о равенстве друг другу соответствующих коэффициентов регрессии.

Таблица 1 Исходные данные

Год	Доход (Y), тыс. руб.
1991	1000
1992	1100
1993	1405
1994	1508
1995	1600
1996	1650
1997	1870
1998	2600
1999	2230
2000	2850
2001	2960

Таблица 2 Исходные данные

Год	Доход (Y), тыс. руб.
2002	2950
2003	3200
2004	3750
2005	4210
2006	4085
2007	5042
2008	5200
2009	5640
2010	5870
2011	6050
2012	6250
2013	6740
2014	6980
2015	7240
2016	7500
2017	7850

Другими словами, будет ли уравнение регрессии одним и тем же для обеих выборок?

1. По МНК оценить коэффициенты линейной регрессии для первой и второй выборок. Оценить коэффициенты регрессии a_1 и a_2 для линейной модели можно с помощью инструмента MS Excel «Регрессия» (меню Данные→Анализ данных (не забыть поставить галочку в пункте Остатки)).

2. Найти суммы квадратов отклонений (остатков)

$$\sum_i \varepsilon_{ik}^2 \ (k = 1{,}2)$$

значений y_i от линий регрессии RSS_1 и RSS_2 соответственно для первого и второго уравнений регрессии. Значения случайных отклонений ε_i представлены в столбце «Остатки». Проведя дополнительные вычисления, рассчитать суммы квадратов отклонений для первой и второй выборок. Результаты расчетов для суммы квадратов остатков представлены на рисунках 1, 2.

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное Y*	*Остатки*	*ε^2*
1	900,77	99,23	9846,05
2	1098,31	1,69	2,86
3	1295,85	109,15	11914,71
4	1493,38	14,62	213,69
5	1690,92	-90,92	8266,12
6	1888,45	-238,45	56860,57
7	2085,99	-215,99	46652,07
8	2283,53	316,47	100154,99
9	2481,06	-251,06	63032,95
10	2678,60	171,40	29377,96
11	2876,14	83,86	7033,11
		RSS1	333355,08

Рисунок 1 Сумма квадратов остатков RSS_1

ВЫВОД ОСТАТКА			
Наблюдение	*Предсказанное Y*	*Остатки*	*ε^2*
1	3106,96	-156,96	24637,46
2	3430,68	-230,68	53211,63
3	3754,39	-4,39	19,27
4	4078,10	131,90	17396,83
5	4401,82	-316,82	100372,49
6	4725,53	316,47	100153,63
7	5049,24	150,76	22727,78
8	5372,96	267,04	71312,56
9	5696,67	173,33	30043,59
10	6020,38	29,62	877,21
11	6344,10	-94,10	8853,98
12	6667,81	72,19	5211,57
13	6991,52	-11,52	132,76
14	7315,24	-75,24	5660,35
15	7638,95	-138,95	19306,69
16	7962,66	-112,66	12692,67
		RSS2	472610,48

Рисунок 2 Сумма квадратов остатков RSS_2

3. По объединенной выборке объема $(n_1 + n_2)$ необходимо построить еще одно уравнение регрессии, для которого сумма квадратов отклонений y_i от уравнения регрессии равна RSS (рисунок 3).

ВЫВОД ОСТАТКА			
Наблюдение	*Предсказанное Y*	*Остатки*	*ε^2*
1	456,63	543,37	295248,48
2	732,99	367,01	134697,66
3	1009,34	395,66	156543,58
4	1285,70	222,30	49417,27
5	1562,06	37,94	1439,75
6	1838,41	-188,41	35499,04
7	2114,77	-244,77	59911,28
8	2391,12	208,88	43629,30
9	2667,48	-437,48	191388,44
10	2943,84	-93,84	8805,11
11	3220,19	-260,19	67699,61
12	3496,55	-546,55	298714,08
13	3772,90	-572,90	328218,23
14	4049,26	-299,26	89556,10
15	4325,62	-115,62	13366,87
16	4601,97	-516,97	267259,12
17	4878,33	163,67	26788,84
18	5154,68	45,32	2053,64
19	5431,04	208,96	43664,75
20	5707,39	162,61	26440,45
21	5983,75	66,25	4388,97
22	6260,11	-10,11	102,14
23	6536,46	203,54	41427,49
24	6812,82	167,18	27949,66
25	7089,17	150,83	22748,36
26	7365,53	134,47	18082,09
27	7641,89	208,11	43311,34
		RSS	2298351,68

Рисунок 3 Сумма квадратов остатков RSS_2

4. Рассчитаем тестовую статистику Чоу:

$$F_{fact} = \frac{(2298351{,}68 - 333355{,}08 - 472610{,}48)/2}{(333355{,}08 - 472610{,}48)/(11 + 16 - 2 \cdot 2)} = 21{,}29.$$

Критическое значение F – статистики можно найти с помощью функции **FРАСПОБР**. Получаем $F_{tabl} = 3{,}42$. Так как $F_{fact} > F_{tabl}$, то нулевая гипотеза отклоняется.

Приведенный выше анализ важен для ответа на вопрос, можно ли за весь рассматриваемый период построить единое уравнение регрессии (рисунок 4 а) или нужно разбить временной интервал на части и на каждой из них построить свое уравнение регрессии (рисунок 4 б)

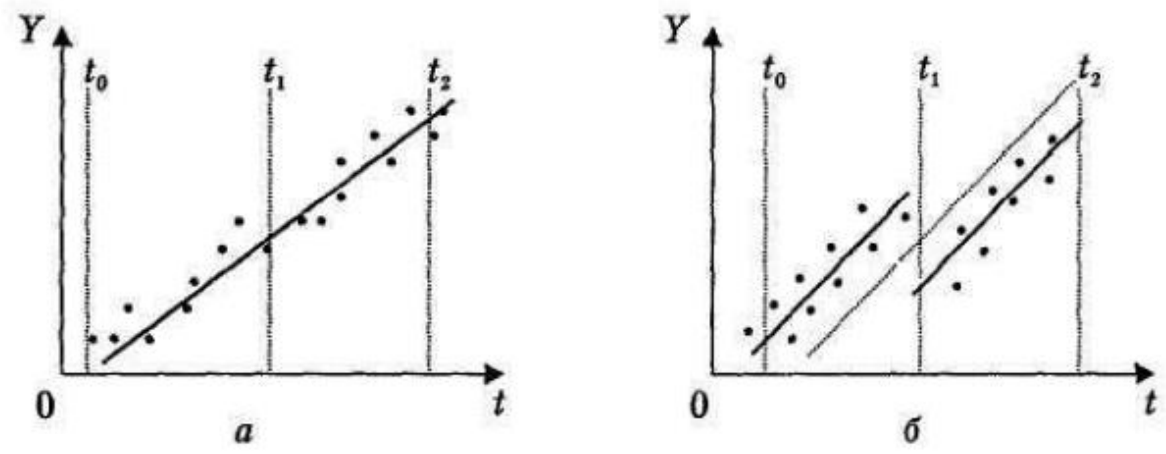

Рисунок 4 Графики моделей линии регрессии

Таблица 3 Задания по вариантам

Вариант №1		Вариант №2		Вариант №3		Вариант №4		Вариант №5		Вариант №6		Вариант №7		Вариант №8		Вариант №9		Вариант №10	
	y		y		y		y		y		y		y		y		y		y
1-ая выборка	172	1-ая выборка	67,9	1-ая выборка	342,7	1-ая выборка	252,7	1-ая выборка	208,7	1-ая выборка	241,7	1-ая выборка	73,7	1-ая выборка	352,7	1-ая выборка	151,7	1-ая выборка	70,4
	132		58,6		296,5		281,5		167,5		213,5		65,5		246,5		173,5		70,3
	372		92,5		348,2		168,2		174,2		218,2		68,2		348,2		149,2		75,7
	182		14,8		344,5		254,5		170,5		216,5		68,5		344,5		146,5		38,9
	137		28,4		193,3		253,3		193,3		215,3		67,3		343,3		145,3		58,8
	232		90,8		345,8		305,8		171,8		227,8		69,8		345,8		147,8		47,9
	170		88,2		343		253		169		215		71		383		145		11,8
	242		48,8		347,1		257,1		173,1		219,1		71,1		347,1		149,1		89,1
	120		21,2		371,2		251,2		167,2		213,2		68,2		341,2		143,2		87,3
	502		16,5		350,4		290,4		176,4		228,4		80,4		380,4		158,4		55,1
2-ая выборка	252		93,5		354,9		254,9		179,9		216,9	2-ая выборка	76,9		354,9		166,9		29,5
	235	2-ая выборка	24,2	2-ая выборка	345,2		255,2		171,2		217,2		69,2	2-ая выборка	245,2		152,2		38,7
	452		66,7		351,5		381,5	2-ая выборка	177,5	2-ая выборка	279,5		81,5		351,5		199,5	2-ая выборка	90,5
	372		18,0		399,4	2-ая выборка	259,4		205,4		221,4		89,4		349,4	2-ая выборка	201,4		11,8
	292		12,0		346,5		256,5		172,5		218,5		70,5		346,5		208,5		62,3
	342		22		386,5		281,5		217,5		277,4		94,4		401,4		227,4		20,8
	392		32		426,5		306,5		262,5		274,5		75,5		398,5		234,5		71,3
	442		42		466,5		331,5		307,5		333,4		99,4		453,4		253,4		29,8
	497		52		506,5		356,5		352,5		330,5		80,5		450,5		260,5		80,3
	547		62		546,5		381,5		397,5		389,4		104		505,4		279,4		38,8
	605		72		586,5		406,5		442,5		386,5		85,5		502,5		286,5		89,3
	655		82		626,5		431,5		487,5		445,4		109		557,4		305,4		47,8
	705		92		666,5		456,5		532,5		442,5		90,5		554,5		312,5		98,3
	760		102		706,5		481,5		577,5		501,4		114		609,4		331,4		56,8
	810		112		746,5		506,5		622,5		498,5		95,5		606,5		338,5		107,3
	860		122		786,5		531,5		667,5		557,4		119		661,4		357,4		65,8
	910		132		826,5		556,5		712,5		554,5		101		658,5		364,5		116,3
	960		142		866,5		581,5		757,5		613,4		124		713,4		383,4		74,8