

МУЛЬТИКОЛЛИНЕАРНОСТЬ

КРАТКОЕ СОДЕРЖАНИЕ
ПРЕДЫДУЩЕЙ ПАРЫ...

НАПОМИНАЮ...

- Мультиколлинеарность - это взаимное влияние факторов друг на друга
- Т.е. модель начинает отражать не только влияние фактора X_i на переменную Y , но и влияние фактора X_i на фактор X_j
- Это смещает коэффициент корреляции и стандартные ошибки -> модель перестает быть информативной и не может быть использована для анализа

СТРОГАЯ МУЛЬТИКОЛЛИНЕАРНОСТЬ

- Это наличие функциональной связи между факторами
- Как лечится? Исключением одного из факторов

Задача 2.2.1

	A	B	C	D	E	F	G
1		y	x1	x2			
2	1	21	11	42			
3	2	27	15	57,75			
4	3	30	16	60			
5	4	45	25	93,75			
6	5	58	32	120,8			
7	6	60	34	127,5			
8	7	69	39	144,8			
9	8	75	41	153,8			
10	9	92	51	192			
11	10	105	58	219			
12	11	114	63	236,3			
13	12	123	69	257,3			
14	13	138	76	285			
15	14	139	77	288,8			
16	15	141	78	293,3			
17	16	135	75	281,3			
18	17	146	81	303,8			
19	18	149	83	310,5			
20	19	158	87	326,3			
21	20	167	92	346,5			

1. Вводим данные задачи

2. Нужно проверить, не вырождена ли матрица $(X'X)$

3. Для этого найдем X' (функция ТРАНСП)

4. Найдем произведение $X'X$ (функция МУМНОЖ)

	y	x1	x2		11,2	15,4	16	25	32,2	34	38,6	41	51,2	58,4	63	68,6	76	77	78,2	75	81	82,8	87	92,4
1	21	11	42		42	57,75	60	93,75	120,75	127,5	144,75	153,75	192	219	236,25	257,25	285	288,75	293,25	281,25	303,75	310,5	326,25	346,5
2	27	15	57,75																					
3	30	16	60		74283	278562																		
4	45	25	93,75		278562	1044608																		
5	58	32	120,8																					
6	60	34	127,5																					
7	69	39	144,8																					
8	75	41	153,8																					
9	92	51	192																					
10	105	58	219																					
11	114	63	236,3																					
12	123	69	257,3																					
13	138	76	285																					
14	139	77	288,8																					
15	141	78	293,3																					
16	135	75	281,3																					
17	146	81	303,8																					
18	149	83	310,5																					
19	158	87	326,3																					
20	167	92	346,5																					

Матрица будет вырожденной, если ее определитель равен 0
Найдем определитель полученной матрицы с помощью функции
МОПРЕД()

	y	x1	x2		11,2	15,4	16	25	32,2	34	38,6	41	51,2	58,4	63	68,6	76	77	78,2	75	81	82,8	87	92,4	
1	21	11	42		42	57,75	60	93,75	120,75	127,5	144,75	153,75	192	219	236,25	257,25	285	288,75	293,25	281,25	303,75	310,5	326,25	346,5	
2	27	15	57,75																						
3	30	16	60		74283	278562																			
4	45	25	93,75		278562	1044608																			
5	58	32	120,8																						
6	60	34	127,5		0																				
7	69	39	144,8																						
8	75	41	153,8																						
9	92	51	192																						
10	105	58	219																						
11	114	63	236,3																						
12	123	69	257,3																						
13	138	76	285																						
14	139	77	288,8																						
15	141	78	293,3																						
16	135	75	281,3																						
17	146	81	303,8																						
18	149	83	310,5																						
19	158	87	326,3																						
20	167	92	346,5																						

Есть подозрения...

Вызовем анализ данных. В регрессии выберем Y и все X, которые даны по условию

74283	278562	Вывод итогов								
278562	1044608									
<i>Регрессионная статистика</i>										
0		Множ	0,9999							
		R-квад	0,9998							
		Норми	0,9442							
		Станд	0,7067							
		Наблк	20							
<i>Дисперсионный анализ</i>										
			<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>значение F</i>			
		Регресс	2	43488	21744	87064	0			
		Остаток	18	8,991	0,4995					
		Итого	20	43497						
<i>Коэффициентная статистика</i>										
			<i>коэффициент</i>	<i>станд. ошибка</i>	<i>t-статистика</i>	<i>значимость</i>	<i>нижние 95%</i>	<i>верхние 95,0%</i>		
		Y-пере	-0,057	0,373	-0,152	0,8811	-0,8	0,73	-0,84	0,7269
		Перемен	0	0	65535	#####	0	0	0	0
		Перемен	0,4814	0,002	295,07	1E-34	0,48	0,48	0,478	0,4849

Ничего не посчиталось для одного из X

Т.е. он является причиной СТРОГОЙ мультиколлинеарности.

Надо его исключить

НЕСТРОГАЯ МУЛЬТИКОЛЛИНЕАРНОСТЬ

- Это наличие корреляционной связи между факторами (связь есть, но она опосредованная)
- Матрица $X'X$ близка к вырождению
- Разрешается ридж-регрессией

Задача 2.2.2

	y	x1	x2
1	2,80	4,40	3,00
2	2,30	3,68	2,51
3	2,50	3,92	2,67
4	2,70	4,32	2,94
5	2,60	4,24	2,89
6	2,40	3,76	2,56
7	3,50	5,52	3,76
8	2,10	3,36	2,29
9	1,90	3,04	2,07
10	3,80	5,92	4,03
11	2,70	4,24	2,89
12	2,90	4,72	3,22
13	3,30	5,28	3,60
14	2,70	4,24	2,89
15	1,80	3,04	2,07
16	1,90	3,12	2,13
17	2,20	3,60	2,46
18	2,30	3,68	2,51
19	3,60	5,76	3,92
20	2,20	3,60	2,45

1. Вводим данные

2. Найдем X' (функция ТРАНСП)

3. Найдем произведение матриц X' и X
(функция МУМНОЖ)

	y	x1	x2		4,40	3,68	3,92	4,32	4,24	3,76	5,52	3,36	3,04	5,92	4,24	4,72	5,28	4,24	3,04	3,12	3,60	3,68	5,76	3,60
1	2,80	4,40	3,00		3,00	2,51	2,67	2,94	2,89	2,56	3,76	2,29	2,07	4,03	2,89	3,22	3,60	2,89	2,07	2,13	2,46	2,51	3,92	2,45
2	2,30	3,68	2,51																					
3	2,50	3,92	2,67		362,82	247,23																		
4	2,70	4,32	2,94		247,23	168,46																		
5	2,60	4,24	2,89																					
6	2,40	3,76	2,56		0,0633																			
7	3,50	5,52	3,76																					
8	2,10	3,36	2,29																					
9	1,90	3,04	2,07																					
10	3,80	5,92	4,03																					
11	2,70	4,24	2,89																					
12	2,90	4,72	3,22																					
13	3,30	5,28	3,60																					
14	2,70	4,24	2,89																					
15	1,80	3,04	2,07																					
16	1,90	3,12	2,13																					
17	2,20	3,60	2,46																					
18	2,30	3,68	2,51																					
19	3,60	5,76	3,92																					
20	2,20	3,60	2,45																					

Определитель округляется в 0, т.е.
 можем подозревать НЕСТРОГУЮ мультиколлинеарность

НО! Сначала оценим столбец из Анализа данных

362,82	247,23	ВЫВОД ИТОГОВ	
247,23	168,46		
<i>Регрессионная статистика</i>			
0,0633		Множественный R	1
		R-квадрат	1
		Нормированный R-квадрат	1
		Стандартная ошибка	0
		Наблюдения	20
<i>Дисперсионный анализ</i>			
			df SS MS F
		Регрессия	2 6,3 3,1 1567,3 4048 0
		Остаток	17 0 0
		Итого	19 6,3
<i>Коэффициенты</i>			
			П-Значение
		Y-пересечение	-0,1 0,1 -1,9 0,069992193
		Переменная X 1	2,7 2,4 1,1 0,29320898
		Переменная X 2	-2,9 3,6 -0,8 0,424942103
			нижние 95,0%
			верхние 95,0%

Должны совпасть ДВА признака

Если есть что-то одно, то НЕСТРОГАЯ мультиколлинеарность есть,
 Но анализу задачи она не мешает 😊

0	4,32	2,94		83,44	362,82	247							
0	4,24	2,89		56,86	247,23	168							
0	3,76	2,56		0,1		0,1	0	0					
0	5,52	3,76				0	0,1	0					
0	3,36	2,29				0	0	0,1					
0	3,04	2,07	X'X+0,1I	20,1	83,44	57							
0	5,92	4,03		83,44	362,92	247							
0	4,24	2,89		56,86	247,23	169							
0	4,72	3,22											
0	5,28	3,60	(X'X+0,1I)-1	1,0946	-0,152	-0,1							
0	4,24	2,89		-0,152	3,1894	-4,6							
0	3,04	2,07		-0,147	-4,627	6,8							
0	3,12	2,13											
0	3,60	2,46	X'Y	52,2						b0.1	-0,0937		
0	3,68	2,45		227,39							0,4454		
0	5,76	3,92		154,94							0,2975		
0	3,60	2,45											

1. Найдем обратную матрицу (функция МОБР)
2. Найдем произведение X' на Y
3. Получим коэффициенты b , умножив обратную матрицу на $X'Y$

	y	x0	x1	x2	y^0,1	y-y^0,1		1,00	1,00	1,00	1
1	2,80	1,00	4,40	3,00	2,76	0,0017		4,40	3,68	3,92	4
2	2,30	1,00	3,68	2,51	2,29	0,0001		3	2,51	2,7	
3	2,50	1,00	3,92	2,67	2,45	0,0028	X'X	20	83,44	57	
4	2,70	1,00	4,32	2,94	2,71	0,0000		83,44	362,82	247	
5	2,60	1,00	4,24	2,89	2,65	0,0030		56,86	247,23	168	
6	2,40	1,00	3,76	2,56	2,34	0,0033		0,1		0,1	
7	3,50	1,00	5,52	3,76	3,48	0,0003				0	
8	2,10	1,00	3,36	2,29	2,08	0,0002				0	
9	1,90	1,00	3,04	2,07	1,88	0,0006	X'X+0,1I	20,1	83,44	57	
10	3,80	1,00	5,92	4,03	3,74	0,0033		83,44	362,92	247	
11	2,70	1,00	4,24	2,89	2,65	0,0020		56,86	247,23	169	
12	2,90	1,00	4,72	3,22	2,97	0,0045					
13	3,30	1,00	5,28	3,60	3,33	0,0009	(X'X+0,1I)-1	1,0946	-0,152	-0,1	
14	2,70	1,00	4,24	2,89	2,65	0,0020		-0,152	3,1894	-4,6	
15	1,80	1,00	3,04	2,07	1,88	0,0058		-0,147	-4,627	6,8	
16	1,90	1,00	3,12	2,13	1,93	0,0009					
17	2,20	1,00	3,60	2,46	2,24	0,0017	X'Y	52,2			
18	2,30	1,00	3,68	2,51	2,29	0,0001		227,39			
19	3,60	1,00	5,76	3,92	3,64	0,0015		154,94			
20	2,20	1,00	3,60	2,45	2,24	0,0015					
						e^2					

Найдем квадраты отклонений расчетных значений от фактических

y	x0	x1	x2	y^0,1	y-y^0,1		1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2,80	1,00	4,40	3,00	2,76	0,0017		4,40	3,68	3,92	4,32	4,24	3,76	5,52	3,36	3,04	5,92	4,24	4,72	5,28	4
2,30	1,00	3,68	2,51	2,29	0,0001		3	2,51	2,7	2,9	2,89	2,6	3,8	2,29	2,1	4	2,9	3,2	3,6	
2,50	1,00	3,92	2,67	2,45	0,0028	X'X	20	83,44	57											
2,70	1,00	4,32	2,94	2,71	0,0000		83,44	362,82	247											
2,60	1,00	4,24	2,89	2,65	0,0030		56,86	247,23	168											
2,40	1,00	3,76	2,56	2,34	0,0033		0,1		0,1	0	0									
3,50	1,00	5,52	3,76	3,48	0,0003				0	0,1	0									
2,10	1,00	3,36	2,29	2,08	0,0002				0	0	0,1									
1,90	1,00	3,04	2,07	1,88	0,0006	X'X+0,1I	20,1	83,44	57											
3,80	1,00	5,92	4,03	3,74	0,0033		83,44	362,92	247											
2,70	1,00	4,24	2,89	2,65	0,0020		56,86	247,23	169											
2,90	1,00	4,72	3,22	2,97	0,0045															
3,30	1,00	5,28	3,60	3,33	0,0009	(X'X+0,1I)-1	1,0946	-0,152	-0,1											
2,70	1,00	4,24	2,89	2,65	0,0020		-0,152	3,1894	-4,6											
1,80	1,00	3,04	2,07	1,88	0,0058		-0,147	-4,627	6,8											
1,90	1,00	3,12	2,13	1,93	0,0009															
2,20	1,00	3,60	2,46	2,24	0,0017	X'Y	52,2					b0.1	-0,0937	=сумм(G2:G21)						
2,30	1,00	3,68	2,51	2,29	0,0001		227,39						0,4454							
3,60	1,00	5,76	3,92	3,64	0,0015		154,94						0,2975							
2,20	1,00	3,60	2,45	2,24	0,0015															
					e^2															

Найдем сумму квадратов отклонений

	56,86	247,23	168																
	0,1		0,1	0	0														
			0	0,1	0														
			0	0	0,1														
$X'X+0,1I$	20,1	83,44	57																
	83,44	362,92	247																
	56,86	247,23	169																
$(X'X+0,1I)^{-1}$	1,0946	-0,152	-0,1																
	-0,152	3,1894	-4,6																
	-0,147	-4,627	6,8																
$X'Y$	52,2					$b_{0.1}$	-0,0937	0,0362	сумма квадратов отклонений										
	227,39						0,4454												
	154,94						0,2975		остаточная дисперсия для 0,1										
									$=R18/(20-2-1)$										

Вычислим остаточную дисперсию

0006	$X'X+0,1I$	20,1	83,44	57															
0033		83,44	362,92	247															
0020		56,86	247,23	169															
0045																			
0009	$(X'X+0,1I)^{-1}$	1,0946	-0,152	-0,1	$S_{0,1}$	=КОРЕНЬ(14*R21)													
0020		-0,152	3,1894	-4,6		КОРЕНЬ(число)													
0058		-0,147	-4,627	6,8															
0009																			
0017	$X'Y$	52,2			$b_{0,1}$	-0,0937	0,0362	сумма квадратов отклонений											
0001		227,39				0,4454													
0015		154,94				0,2975	остаточная дисперсия для 0,1												
0015							0,002												
2																			

Найдем стандартные ошибки для альфа = 0,1

	83,44	362,82	247																	
	56,86	247,23	168																	
	0,1		0,1	0	0															
			0	0,1	0															
			0	0	0,1															
11	20,1	83,44	57																	
	83,44	362,92	247																	
	56,86	247,23	169																	
11)-1	1,0946	-0,152	-0,1			S0,1	0,0483													
	-0,152	3,1894	-4,6				0,0824													
	-0,147	-4,627	6,8				0,1207													
	52,2					b0,1	-0,0937			0,0362	сумма квадратов отклонений									
	227,39						0,4454													
	154,94						0,2975			остаточная дисперсия для 0,1										
										0,002										

Далее делаем все то же самое для альфа = 0,2 , 0,3 , 0,4.

Из четырех столбцов стандартных ошибок находим столбец с наименьшими значениями

Для наименьших стандартных ошибок вычисляем Стьюдента

ЧТО ДЕЛАЕМ ДАЛЬШЕ?

- Делаете задачу 2.2.1
- Делаете задачу 2.2.2
- Делаете задачу 2.3.1 (определить тип мультиколлинеарности, избавиться от нее, если необходимо или решить задачу множественной регрессии, если она не мешает анализу)
- Делаете задачу 2.3.2 (определить тип мультиколлинеарности, избавиться от нее, если необходимо или решить задачу множественной регрессии, если она не мешает анализу)