Лабораторна робота № 5

Мета роботи: навчитися перевіряти наявність мультиколінеарності за алгоритмом Фаррара-Глобера засобами табличного редактора Microsoft Excel.

Приклад виконання лабораторної роботи

Завдання. За допомогою табличного редактора Microsoft Excel дослідити наявність мультиколінеарності між незалежними змінними (фондовіддачею, продуктивністю праці, питомими інвестиціями), скориставшись алгоритмом Фаррара-Глобера (табл. 5.1). Необхідно:

1. Виконати ідентифікацію змінних.
2. Виконати нормалізацію змінних.
3. Побудувати кореляційну матрицю незалежних змінних r.
4. Перевірити мультиколінеарність незалежних змінних.
   4.1. Усього масиву незалежних змінних за χ^2 − критерієм.
   4.2. Кожної незалежної змінної з рештою змінних за F− критерієм Фішера.
   4.3. Кожної пари незалежних змінних за t −критерієм Стьюдента.
5. Зробити висновки.

Таблиця 5.1 – Вихідні дані

Місяць	Прибуток на місяць, грн.	Фондовіддача, грн.	Продуктивність праці, грн.	Питомі інвестиції, грн.
1	40	12	5	15
2	45	17	7	18
3	40	13	6	16
4	43	14	7	17
5	48	16	6	20
6	39	15	5	15
7	42	14	6	16
8	45	17	9	18
9	38	12	5	19
10	48	18	10	20
11	50	20	11	22
12	48	17	10	21
13	49	18	12	21
14	45	19	8	20
15	49	20	9	22
16	52	22	14	23
17	54	24	15	24
18	51	21	13	20
19	55	25	16	24
20	56	27	18	25

Розв'язання.

1. Ідентифікація змінних та специфікація моделі.

Ідентифікуємо змінні економетричної моделі:

Y − вектор прибутку на місяць (залежна змінна);

X_1 − вектор фондовіддачі (незалежна змінна);

X_2 − вектор продуктивності праці (незалежна змінна);

X_3 − вектор питомих інвестицій (незалежна змінна);

u − вектор залишків (стохастична складова).

Загальний вигляд економетричної моделі:

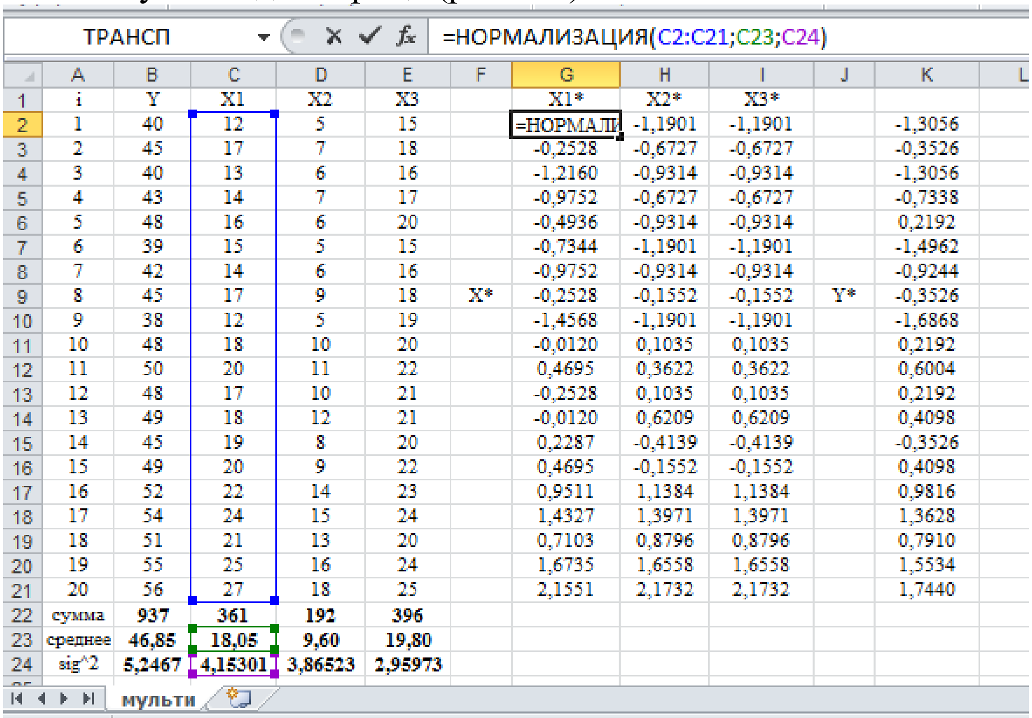
2. Нормалізація змінних.

Обчислимо середні значення та стандартні відхилення залежної змінної Y і незалежних змінних X_1, X_2, X_3, використовуючи функції СРЗНАЧ та СТАНДОТКЛОНП і запишемо в табл. 5.2.

Таблиця 5.2 – Розрахункова таблиця

Місяць	Прибуток на місяць, грн.	Фондовіддача, грн.	Продуктивність праці, грн.	Питомі інвестиції, грн.
1	40	12	5	15
2	45	17	7	18
3	40	13	6	16
4	43	14	7	17
5	48	16	6	20
6	39	15	5	15
7	42	14	6	16
8	45	17	9	18
9	38	12	5	19
10	48	18	10	20
11	50	20	11	22
12	48	17	10	21
13	49	18	12	21
14	45	19	8	20
15	49	20	9	22
16	52	22	14	23
17	54	24	15	24
18	51	21	13	20
19	55	25	16	24
20	56	27	18	25
сума	**937**	**361**	**192**	**396**
середнє	**46,85**	**18,05**	**9,60**	**19,80**
σ^2	**5,24667**	**4,15301**	**3,86523**	**2,95973**

За допомогою статистичної функції НОРМАЛИЗАЦИЯ визначимо значення нормалізованих незалежних змінних X_1^*, X_2^*, X_3^*, залежної змінної Y^* та запишемо у вигляді матриць (рис. 5.1).



	A	B	C	D	E	F	G	H	I	J	K
1	i	Y	X1	X2	X3		X1*	X2*	X3*		
2	1	40	12	5	15		=НОРМАЛИ	-1,1901	-1,1901		-1,3056
3	2	45	17	7	18		-0,2528	-0,6727	-0,6727		-0,3526
4	3	40	13	6	16		-1,2160	-0,9314	-0,9314		-1,3056
5	4	43	14	7	17		-0,9752	-0,6727	-0,6727		-0,7338
6	5	48	16	6	20		-0,4936	-0,9314	-0,9314		0,2192
7	6	39	15	5	15		-0,7344	-1,1901	-1,1901		-1,4962
8	7	42	14	6	16		-0,9752	-0,9314	-0,9314		-0,9244
9	8	45	17	9	18	X*	-0,2528	-0,1552	-0,1552	Y*	-0,3526
10	9	38	12	5	19		-1,4568	-1,1901	-1,1901		-1,6868
11	10	48	18	10	20		-0,0120	0,1035	0,1035		0,2192
12	11	50	20	11	22		0,4695	0,3622	0,3622		0,6004
13	12	48	17	10	21		-0,2528	0,1035	0,1035		0,2192
14	13	49	18	12	21		-0,0120	0,6209	0,6209		0,4098
15	14	45	19	8	20		0,2287	-0,4139	-0,4139		-0,3526
16	15	49	20	9	22		0,4695	-0,1552	-0,1552		0,4098
17	16	52	22	14	23		0,9511	1,1384	1,1384		0,9816
18	17	54	24	15	24		1,4327	1,3971	1,3971		1,3628
19	18	51	21	13	20		0,7103	0,8796	0,8796		0,7910
20	19	55	25	16	24		1,6735	1,6558	1,6558		1,5534
21	20	56	27	18	25		2,1551	2,1732	2,1732		1,7440
22	сумма	937	361	192	396						
23	среднее	46,85	18,05	9,60	19,80						
24	sig^2	5,2467	4,15301	3,86523	2,95973						

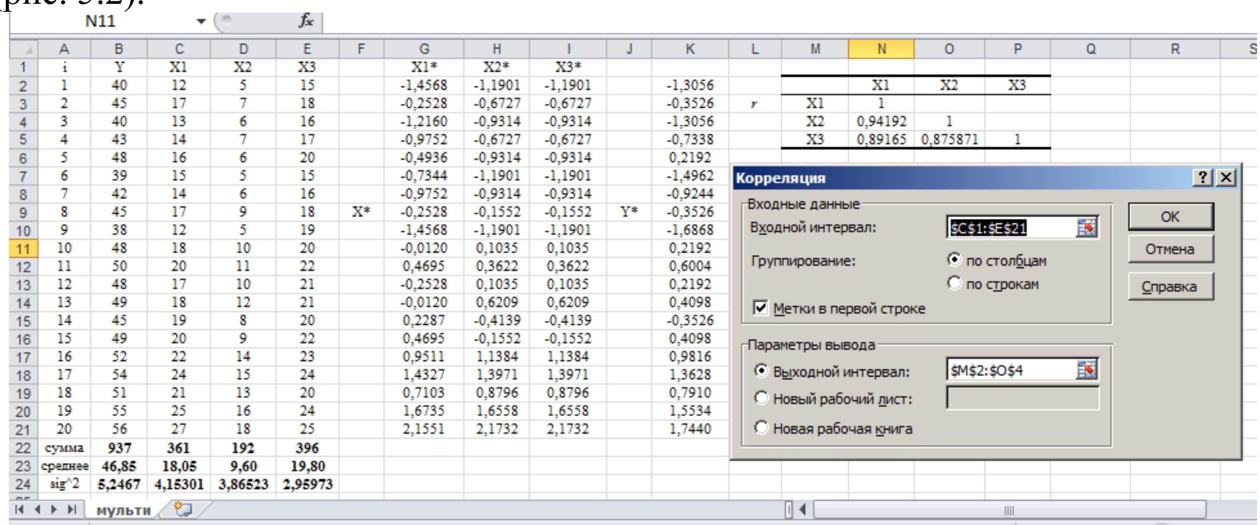
Рисунок 5.1 – Нормалізація змінних

3. *Побудова кореляційної матриці r.*

Побудуємо кореляційну матрицю незалежних змінних

$$r = \frac{1}{n} X^{*T} X^{*} = \begin{pmatrix} r_{X_1 X_1} & r_{X_1 X_2} & \dots & r_{X_1 X_k} \\ r_{X_2 X_1} & r_{X_2 X_2} & \dots & r_{X_2 X_k} \\ \dots & \dots & \dots & \dots \\ r_{X_k X_1} & r_{X_k X_2} & \dots & r_{X_k X_k} \end{pmatrix}$$

за допомогою інструменту ДАННЫЕ → АНАЛИЗ ДАННЫХ → КОРРЕЛЯЦИЯ (рис. 5.2).



	A	B	C	D	E	F	G	H	I	J	K
1	i	Y	X1	X2	X3		X1*	X2*	X3*		
2	1	40	12	5	15		-1,4568	-1,1901	-1,1901		-1,3056
3	2	45	17	7	18		-0,2528	-0,6727	-0,6727		-0,3526
4	3	40	13	6	16		-1,2160	-0,9314	-0,9314		-1,3056
5	4	43	14	7	17		-0,9752	-0,6727	-0,6727		-0,7338
6	5	48	16	6	20		-0,4936	-0,9314	-0,9314		0,2192
7	6	39	15	5	15		-0,7344	-1,1901	-1,1901		-1,4962
8	7	42	14	6	16		-0,9752	-0,9314	-0,9314		-0,9244
9	8	45	17	9	18	X*	-0,2528	-0,1552	-0,1552	Y*	-0,3526
10	9	38	12	5	19		-1,4568	-1,1901	-1,1901		-1,6868
11	10	48	18	10	20		-0,0120	0,1035	0,1035		0,2192
12	11	50	20	11	22		0,4695	0,3622	0,3622		0,6004
13	12	48	17	10	21		-0,2528	0,1035	0,1035		0,2192
14	13	49	18	12	21		-0,0120	0,6209	0,6209		0,4098
15	14	45	19	8	20		0,2287	-0,4139	-0,4139		-0,3526
16	15	49	20	9	22		0,4695	-0,1552	-0,1552		0,4098
17	16	52	22	14	23		0,9511	1,1384	1,1384		0,9816
18	17	54	24	15	24		1,4327	1,3971	1,3971		1,3628
19	18	51	21	13	20		0,7103	0,8796	0,8796		0,7910
20	19	55	25	16	24		1,6735	1,6558	1,6558		1,5534
21	20	56	27	18	25		2,1551	2,1732	2,1732		1,7440
22	сумма	937	361	192	396						
23	среднее	46,85	18,05	9,60	19,80						
24	sig^2	5,2467	4,15301	3,86523	2,95973						

L	M	N	O	P
		X1	X2	X3
r	X1	1		
	X2	0,94192	1	
	X3	0,89165	0,875871	1

Рисунок 5.2 – Побудова кореляційної матриці

Будемо мати матрицю:

$$r=\begin{Vmatrix} 1 & & \\ 0{,}9419187 & 1 & \\ 0{,}8916521 & 0{,}87587145 & 1 \end{Vmatrix}$$

Якщо діагональні елементи кореляційної матриці не дорівнюють одиниці, то на діагоналі цієї матриці ставимо одиниці, а до інших елементів додаємо різницю між одиницею і значенням діагонального елемента.

Кожен елемент цієї матриці характеризує тісноту зв'язку однієї незалежної змінної з іншою. Оскільки діагональні елементи характеризують тісноту зв'язку кожної незалежної змінної з цією самою змінною, то вони дорівнюють одиниці.

Інші елементи кореляційної матриці *r* є парними коефіцієнтами кореляції незалежних змінних. На основі цих коефіцієнтів можна зробити висновок, що між змінними X_1, X_2, X_3 існує сильний зв'язок (додаток А).

4. Перевірка мультиколінеарності незалежних змінних.

Мультиколінеарність – це існування тісної лінійної залежності (сильної кореляції) між двома чи більше незалежними змінними моделі.

4.1. Перевірка мультиколінеарності усього масиву незалежних змінних за χ^2 –критерієм.

За допомогою функції МОПРЕД визначимо значення визначника матриці *r*:

$$\det(r) = \quad 0{,}022$$

Оскільки $\det(r) \to 0$, то в масиві незалежних змінних може існувати мультиколінеарність (чим ближче визначник до нуля, тим сильнішою є мультиколінеарність).

Обчислимо критерій Пірсона χ^2 (хі-квадрат):

$$\chi^2 = -\left(n-1-\frac{2m+5}{6}\right)\ln|r| = \quad 65{,}6610$$

Порівняємо фактичне значення χ^2 − критерію з табличним для $k=\frac{m(m-1)}{2}=3$ – числа ступенів свободи та $\alpha=0{,}05$ – рівня значущості: $\chi^2_{\text{табл}}=7{,}81$ (додаток К). Для знаходження табличного значення окрім спеціальних таблиць можна застосувати функцію ХИ2.ОБР.ПХ(0,05; 3)=7,8147.

Оскільки $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$, то в масиві незалежних змінних (продуктивність праці, питомі інвестиції та фондовіддача) існує мультиколінеарність і необхідно проводити подальше дослідження.

4.2. Перевірка мультиколінеарності кожної незалежної змінної з рештою змінних за F − критерієм Фішера

Визначимо матрицю помилок, тобто матрицю обернену до кореляційної матриці *r* (функція МОБР, рис. 5.3).

M13 | f_x | {=МОБР(N3:P5)}

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	i	Y	X1	X2	X3		X1*	X2*	X3*							
2	1	40	12	5	15		-1,4568	-1,1901	-1,1901		-1,3056			X1	X2	X3
3	2	45	17	7	18		-0,2528	-0,6727	-0,6727		-0,3526		X1	1	0,941919	0,89165
4	3	40	13	6	16		-1,2160	-0,9314	-0,9314		-1,3056	r	X2	0,94192	1	0,87587
5	4	43	14	7	17		-0,9752	-0,6727	-0,6727		-0,7338		X3	0,89165	0,875871	1
6	5	48	16	6	20		-0,4936	-0,9314	-0,9314		0,2192					
7	6	39	15	5	15		-0,7344	-1,1901	-1,1901		-1,4962			k	3	
8	7	42	14	6	16		-0,9752	-0,9314	-0,9314		-0,9244	det(r)	0,022	alfa	0,05	
9	8	45	17	9	18	X*	-0,2528	-0,1552	-0,1552	Y*	-0,3526					
10	9	38	12	5	19		-1,4568	-1,1901	-1,1901		-1,6868	hi^2	65,6610			
11	10	48	18	10	20		-0,0120	0,1035	0,1035		0,2192	табл	7,8147			
12	11	50	20	11	22		0,4695	0,3622	0,3622		0,6004					
13	12	48	17	10	21		-0,2528	0,1035	0,1035		0,2192		10,6712	-7,3760	-3,0546	
14	13	49	18	12	21		-0,0120	0,6209	0,6209		0,4098	r^-1	-7,3760	9,3929	-1,6502	
15	14	45	19	8	20		0,2287	-0,4139	-0,4139		-0,3526		-3,0546	-1,6502	5,1690	
16	15	49	20	9	22		0,4695	-0,1552	-0,1552		0,4098					
17	16	52	22	14	23		0,9511	1,1384	1,1384		0,9816					
18	17	54	24	15	24		1,4327	1,3971	1,3971		1,3628					
19	18	51	21	13	20		0,7103	0,8796	0,8796		0,7910					
20	19	55	25	16	24		1,6735	1,6558	1,6558		1,5534					
21	20	56	27	18	25		2,1551	2,1732	2,1732		1,7440					
22	сумма	**937**	**361**	**192**	**396**											
23	среднее	**46,85**	**18,05**	**9,60**	**19,80**											
24	sig^2	**5,2467**	**4,15301**	**3,86523**	**2,95973**											

мульти

Рисунок 5.3 – Матриця, обернена до кореляційної матриці

Обчислимо значення *F* – критеріїв, застосовуючи знайдену матрицю і формулу $F_k = (c_{kk} - 1)\dfrac{n - m}{m - 1}$:

$$F_1 = (c_{11} - 1)\frac{20 - 3}{3 - 1} = \quad 82{,}2052$$

$$F_2 = (c_{22} - 1)\frac{20 - 3}{3 - 1} = \quad 71{,}3398$$

$$F_3 = (c_{33} - 1)\frac{20 - 3}{3 - 1} = \quad 35{,}4365$$

Обчислені значення *F* − критеріїв порівнюємо з табличним значенням $F_{\text{табл}} = F(\alpha; k_1; k_2) = 3{,}59$ (додаток Е), де $\alpha = 0{,}05$ – рівень значущості, $k_1 = m - 1 = 3 - 1 = 2$, $k_2 = n - m = 20 - 3 = 17$ – ступені свободи. Табличне значення можна також визначити за допомогою функції FРАСПОБР(0,05;2;17)=3,5915.

Якщо *F* − критерій більше табличного значення, це значить, що *k* − та змінна залежить від всіх інших в масиві, тобто необхідно розглядати питання про її виключення з переліку змінних (рис. 5.4).

S20

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	i	Y	X1	X2	X3		X1*	X2*	X3*								
2	1	40	12	5	15		-1,4568	-1,1901	-1,1901		-1,3056			X1	X2	X3	
3	2	45	17	7	18		-0,2528	-0,6727	-0,6727		-0,3526		X1	1	0,941919	0,89165	
4	3	40	13	6	16		-1,2160	-0,9314	-0,9314		-1,3056	r	X2	0,94192	1	0,87587	
5	4	43	14	7	17		-0,9752	-0,6727	-0,6727		-0,7338		X3	0,89165	0,875871	1	
6	5	48	16	6	20		-0,4936	-0,9314	-0,9314		0,2192						
7	6	39	15	5	15		-0,7344	-1,1901	-1,1901		-1,4962			k	3		
8	7	42	14	6	16		-0,9752	-0,9314	-0,9314		-0,9244	det(r)	0,022	alfa	0,05		
9	8	45	17	9	18	X*	-0,2528	-0,1552	-0,1552	Y*	-0,3526						
10	9	38	12	5	19		-1,4568	-1,1901	-1,1901		-1,6868	hi^2	65,6610	k1	2		
11	10	48	18	10	20		-0,0120	0,1035	0,1035		0,2192	табл	7,8147	k2	17		
12	11	50	20	11	22		0,4695	0,3622	0,3622		0,6004						
13	12	48	17	10	21		-0,2528	0,1035	0,1035		0,2192		10,6712	-7,3760	-3,0546		
14	13	49	18	12	21		-0,0120	0,6209	0,6209		0,4098	r^-1	-7,3760	9,3929	-1,6502		
15	14	45	19	8	20		0,2287	-0,4139	-0,4139		-0,3526		-3,0546	-1,6502	5,1690		
16	15	49	20	9	22		0,4695	-0,1552	-0,1552		0,4098						
17	16	52	22	14	23		0,9511	1,1384	1,1384		0,9816	F_1	82,2052			залежіть від інших	
18	17	54	24	15	24		1,4327	1,3971	1,3971		1,3628	F_2	71,3398	F_табл	3,5915	залежіть від інших	
19	18	51	21	13	20		0,7103	0,8796	0,8796		0,7910	F_3	35,4365			залежіть від інших	
20	19	55	25	16	24		1,6735	1,6558	1,6558		1,5534						
21	20	56	27	18	25		2,1551	2,1732	2,1732		1,7440						
22	сумма	**937**	**361**	**192**	**396**												
23	среднее	**46,85**	**18,05**	**9,60**	**19,80**												
24	sig^2	**5,2467**	**4,15301**	**3,86523**	**2,95973**												

мульти

Рисунок 5.4 – Значення F – критеріїв та порівняння їх з табличним значенням

Використовуючи матрицю помилок $C = r^{-1}$, визначимо значення коефіцієнта детермінації $R^2_{X_k} = 1 - \dfrac{1}{c_{kk}}$ для кожної незалежної змінної:

$R^2_{X_1} =$ 0,90629

$R^2_{X_2} =$ 0,89354

$R^2_{X_3} =$ 0,80654

Оскільки коефіцієнти детермінації наближаються до одиниці, то кожна незалежна змінна мультиколінеарна з іншими.

4.3. Перевірка мультиколінеарності кожної пари незалежних змінних за t –критерієм Стьюдента

Частинні коефіцієнти кореляції r, що показують тісноту зв'язку між двома незалежними змінними за умови, що інші змінні не впливають на цей зв'язок обчислюються за формулою:

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} \cdot c_{jj}}}, \quad k = \overline{1, m}, \; j = \overline{1, m},$$

де c_{kj} – елемент матриці C, що заходиться в k-му рядку і j-му стовпці,

c_{kk} і c_{jj} – діагональні елементи матриці C.

В нашому випадку будемо мати:

$r_{12} =$ 0,73674

$r_{23} =$ 0,23683

$r_{13} =$ 0,41129

Запишемо матрицю частинних коефіцієнтів кореляції:

$$r' = \begin{vmatrix} 1 & 0{,}73674 & 0{,}41129 \\ 0{,}73674 & 1 & 0{,}23683 \\ 0{,}41129 & 0{,}23683 & 1 \end{vmatrix}$$

Порівнявши частинні коефіцієнти кореляції з парними, можна помітити, що частинні коефіцієнти значно менше парних.

Враховуючи значення частинних коефіцієнтів кореляції (додаток А), можна зробити висновок, що зв'язок між фондовіддачею та продуктивністю праці є сильним, якщо не враховувати вплив питомих інвестицій ($r_{12} = 0{,}73674$); зв'язок між фондовіддачею та питомими інвестиціями є помірним, якщо не брати до уваги вплив продуктивності праці ($r_{13} = 0{,}41129$). Зв'язок між продуктивністю праці та питомими інвестиціями є слабким, якщо не враховувати фондовіддачу ($r_{23} = 0{,}23683$).

Визначимо значення t −критеріїв Стюдента для визначення попарної мультиколінеарності двох незалежних змінних за формулою:

$$t_{kj} = \frac{|r_{kj}|\sqrt{n-m}}{\sqrt{1-r_{kj}^2}}.$$

Будемо мати:

$t_{12} =$ 4,49236

$t_{23} =$ 1,00505

$t_{13} =$ 1,86041

Обчислені значення t −критеріїв порівнюємо з табличним $t_{табл} = 2{,}11$ при $n - m = 17$ ступенях свободи і рівні значущості $\alpha = 0{,}05$ (додаток Д). Або, використавши функцію СТЬЮДРАСПОБР(0,05;17)=2,109816 (рис. 5.5).

J	K	L	M	N	O	P	Q	R	S	T
	-1,3056			X1	X2	X3				
	-0,3526		X1	1	0,941919	0,89165				
	-1,3056	r	X2	0,94192	1	0,87587				
	-0,7338		X3	0,89165	0,875871	1				
	0,2192									
	-1,4962			k	3					
	-0,9244	det(r)	0,022	alfa	0,05					
Y*	-0,3526									
	-1,6868	hi^2	65,6610	k1	2					
	0,2192	табл	7,8147	k2	17					
	0,6004									
	0,2192		10,6712	-7,3760	-3,0546					
	0,4098	r^-1	-7,3760	9,3929	-1,6502					
	-0,3526		-3,0546	-1,6502	5,1690					
	0,4098									
	0,9816	F_1	82,2052			залежіть від інших				
	1,3628	F_2	71,3398	F_табл	3,5915	залежіть від інших				
	0,7910	F_3	35,4365			залежіть від інших				
	1,5534									
	1,7440	R_1	0,90629	r_12	0,73674	t_12	4,49236	t_табл	1 і 2 мультиколінеарні	
		R_2	0,89354	r_23	0,23683	t_23	1,00505	2,109816	2 і 3 не мультиколінеарні	
		R_3	0,80654	r_13	0,41129	t_13	1,86041		1 і 3 не мультиколінеарні	

Рисунок 5.5 – Значення t − критеріїв та порівняння їх з табличним значенням

Оскільки $t_{12} > t_{\text{табл}}$, то фондовіддача та продуктивність праці є відповідно мультиколінеарними між собою; $t_{13} < t_{\text{табл}}$, тому відповідно фондовіддача та питомі інвестиції не є мультиколінеарними між собою; $t_{23} < t_{\text{табл}}$, тому продуктивність праці та питомі інвестиції не є мультиколінеарними між собою.

5. Висновки.

Дослідження показали, що в заданому масиві даних мультиколінеарність існує між фондовіддачею та продуктивністю праці. Отже, для того, щоб можна було оцінювати параметри моделі за цією інформацією, необхідно звільнитися від мультиколінеарності.

Якщо пара факторів мультиколінеарні, то один з цих факторів залишають, а інший вилучають з моделі. Найчастіше залишають той фактор, який з економічної точки зору більш вагомий. Якщо із економічних міркувань ні одному з них не можна надати перевагу, то залишають той фактор, який має більший коефіцієнт кореляції із залежною змінною, а потім за допомогою класичного методу найменших квадратів будують відповідну економетричну модель. В нашому випадку треба виключити фактор X_1 як менш вагомий з економічної точки зору або як такий що має більший F –критерій. Аналогічний результат отримаємо, якщо обчислити матрицю кореляції незалежних і залежних змінних:

$$r = \begin{Vmatrix} 1 & 0{,}94346 & 0{,}93394 & 0{,}91572 \\ 0{,}94346 & 1 & 0{,}94192 & 0{,}89165 \\ 0{,}93394 & 0{,}94192 & 1 & 0{,}87587 \\ 0{,}91572 & 0{,}89165 & 0{,}87587 & 1 \end{Vmatrix}$$

Будемо мати, що $r_{21} = r_{YX_1}$ має найбільше значення, отже треба виключити фактор X_1.