

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
Кафедра штучного інтелекту

Звіт з виконання завдань
комп'ютерного практикуму № 6
Кореляційно-регресійний аналіз у середовищі STATISTICA
з кредитного модуля «Багатовимірний статистичний аналіз»

Звіт склав
студент гр. КІ-01
Копцов В.О.
Прийняла: *Ірина Джигирей*

Мета роботи. Засвоїти класичну модель факторного аналізу та основні етапи факторного аналізу, опанувати методи факторизування кореляційної матриці, критерії достатності кількості виділених факторів для пояснення кореляційної матриці, методи ортогонального обертання, набути вмінь та досвіду проведення факторного аналізу сукупності об'єктів з використанням методу головних компонент для зниження розмірності простору ознак.

Завдання. За наданими результатами спостережень згідно індивідуального завдання (Додаток А) виконати факторний аналіз. Розрахунки провести у застосунку Statistica.

Варіант №30

	1 x1	2 x2	3 x3	4 x4	5 x5	6 x6	7 x7
1	-18,4	0,7	-1,7	1,1	-2,6	24	1,6
2	-11,3	3,1	-1,6	3	-3,1	9,3	4,7
3	-12,8	3	-2	3,3	-3,8	10,4	4,6
4	-22,5	4,1	-2,7	2,1	-4,8	19,9	2,5
5	-42,8	1,3	-3,8	0,9	-5,9	58,3	0,9
6	-23,5	3,5	-3,3	2,7	-6,1	19,4	2,8
7	-22,7	3,3	-3,5	3,2	-6,7	18,5	3,1
8	-33,5	3,9	-4,3	2,3	-7,8	28,5	2
9	-56,1	-3,9	-5,5	1,2	-9	62,3	1
10	-37,7	3,9	-5	2,5	-9,2	31,5	1,9
11	-33,3	3,5	-5	3,1	-9,7	27,2	2,3
12	-42,9	3,9	-5,8	2,5	-10,7	35,6	1,8
13	-66,1	7,4	-7,1	1,5	-12,1	64,9	1
14	-54,8	4,3	-6,9	2,2	-12,4	47	1,4
15	-45,2	3,7	-6,6	3	-12,7	36,9	1,8
16	-51,2	3,8	-7,2	2,7	-13,5	42	1,6
17	-73,9	5,1	-8,6	1,8	-15	67,1	1
18	-75,8	4,9	-8,9	1,9	-15,7	67,9	1
19	-58,7	3,9	-8,3	2,8	-15,7	48,2	1,4
20	-58,9	3,8	-8,6	2,9	-16,4	48,2	1,4
21	-80,3	4,6	-9,9	2,1	-17,8	69,4	1
22	-102,1	6,6	-11,2	1,6	-19,1	98,2	0,7
23	-74,8	4,1	-10,2	2,6	-18,9	61,9	1,1
24	-66,7	3,9	-10	3	-19,2	54,4	1,3
25	-85,8	4,3	-11,2	2,4	-20,6	72	1

Хід виконання завдань практикуму

1. Для початку нормалізуємо дані.

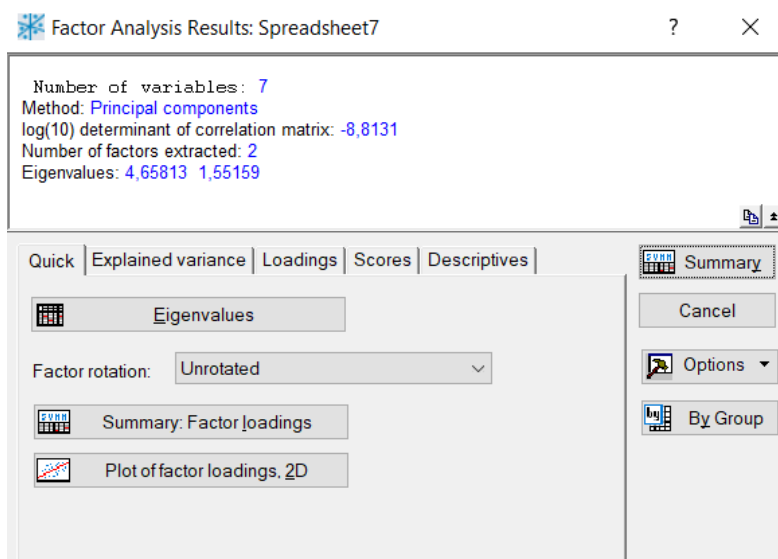
	1 x1	2 x2	3 x3	4 x4	5 x5	6 x6	7 x7
1	1,292	-1,427	1,541	-1,806	1,611	-0,925	-0,185
2	1,581	-0,257	1,574	0,970	1,521	-1,575	2,744
3	1,520	-0,306	1,442	1,409	1,395	-1,527	2,649
4	1,124	0,230	1,210	-0,345	1,215	-1,106	0,665
5	0,297	-1,135	0,846	-2,099	1,016	0,592	-0,846
6	1,084	-0,062	1,011	0,532	0,980	-1,129	0,949
7	1,116	-0,160	0,945	1,263	0,872	-1,168	1,232
8	0,676	0,133	0,680	-0,053	0,674	-0,726	0,193
9	-0,246	-3,669	0,283	-1,660	0,458	0,769	-0,752
10	0,505	0,133	0,449	0,240	0,422	-0,593	0,098
11	0,684	-0,062	0,449	1,117	0,332	-0,784	0,476
12	0,292	0,133	0,184	0,240	0,151	-0,412	0,004
13	-0,654	1,838	-0,246	-1,222	-0,101	0,884	-0,752
14	-0,193	0,327	-0,180	-0,199	-0,155	0,092	-0,374
15	0,199	0,035	-0,081	0,970	-0,209	-0,355	0,004
16	-0,046	0,084	-0,279	0,532	-0,353	-0,129	-0,185
17	-0,972	0,717	-0,743	-0,783	-0,624	0,981	-0,752
18	-1,049	0,620	-0,842	-0,637	-0,750	1,016	-0,752
19	-0,352	0,133	-0,643	0,678	-0,750	0,145	-0,374
20	-0,360	0,084	-0,743	0,824	-0,876	0,145	-0,374
21	-1,233	0,474	-1,173	-0,345	-1,128	1,083	-0,752
22	-2,122	1,448	-1,603	-1,076	-1,362	2,356	-1,035
23	-1,008	0,230	-1,272	0,386	-1,326	0,751	-0,658
24	-0,678	0,133	-1,206	0,970	-1,380	0,419	-0,469
25	-1,457	0,327	-1,603	0,094	-1,633	1,198	-0,752

Матриця кореляцій

Variable	Correlations (Spreadsheet7) Casewise deletion of MD N=25						
	x1	x2	x3	x4	x5	x6	x7
x1	1,00	-0,38	0,96	0,29	0,92	-0,96	0,80
x2	-0,38	1,00	-0,43	0,23	-0,44	0,25	-0,11
x3	0,96	-0,43	1,00	0,04	0,99	-0,86	0,73
x4	0,29	0,23	0,04	1,00	-0,06	-0,50	0,59
x5	0,92	-0,44	0,99	-0,06	1,00	-0,80	0,68
x6	-0,96	0,25	-0,86	-0,50	-0,80	1,00	-0,84
x7	0,80	-0,11	0,73	0,59	0,68	-0,84	1,00

Є дуже сильний зв'язок між x1, x3 і x5. І сильний обернений зв'язок між x1(і відповідно інших змінних прямо пов'язаних з ним) і x6. x7 сильно пов'язаний з x1. x4 має помітні зв'язки з x6 і x7. x2 не має жодних зв'язків.

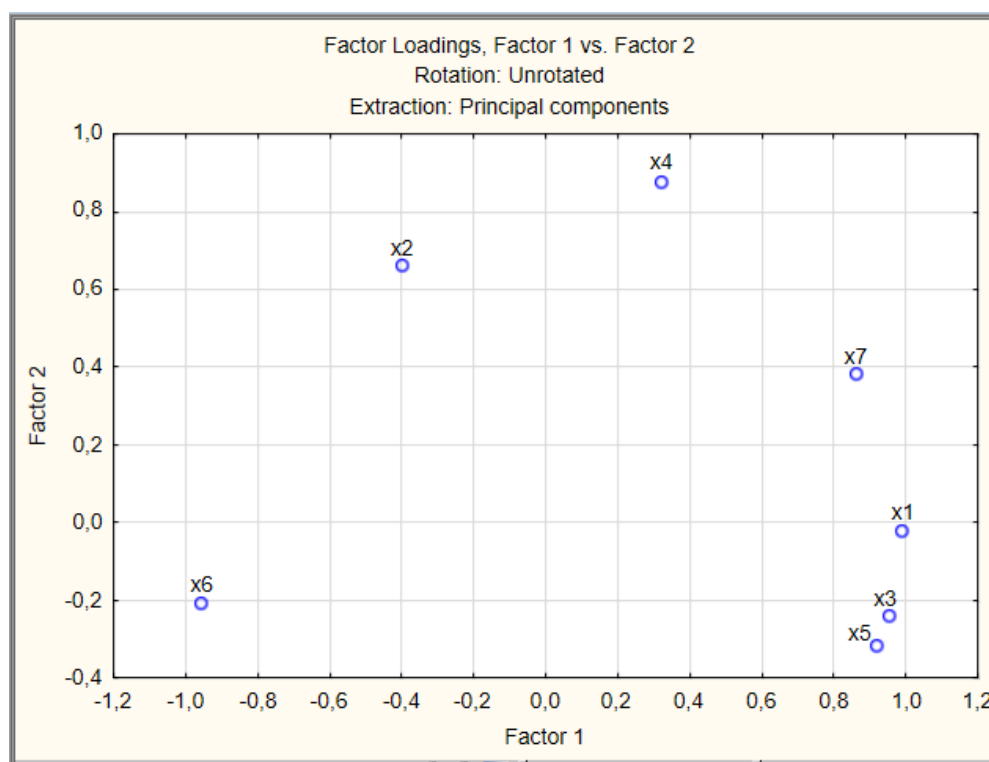
3 семи змінних отримали 2 критерії.



Власні значення(+загальна дисп)

Eigenvalues (Spreadsheet7) Extraction: Principal components				
Value	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	4,658126	66,54465	4,658126	66,54465
2	1,551586	22,16551	6,209711	88,71016

Графік навантажень

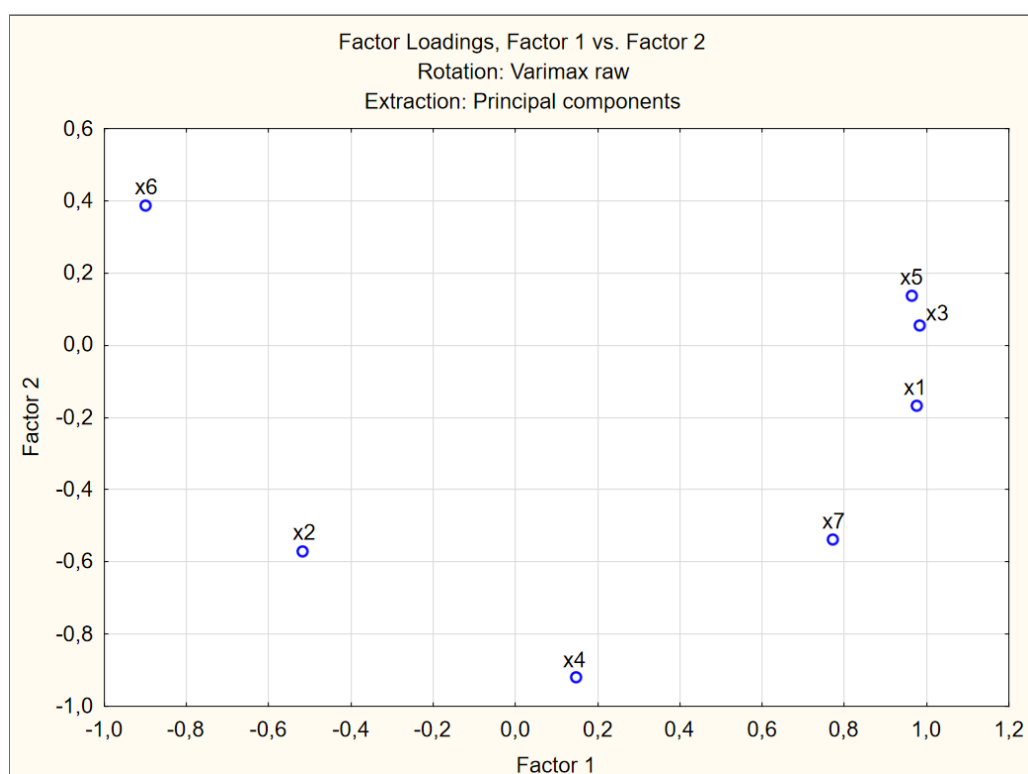


Навантаження факторів

Factor Loadings (Unrotated) (Spreadsheet7)				
Extraction: Principal components				
(Marked loadings are >,700000)				
Variable	Factor 1	Factor 2		
x1	0,990239	-0,021323		
x2	-0,398477	0,659124		
x3	0,954799	-0,242802		
x4	0,321412	0,876104		
x5	0,921027	-0,318411		
x6	-0,955571	-0,207606		
x7	0,861636	0,381693		
Expl.Var	4,658126	1,551586		
Prp.Totl	0,665447	0,221655		

Varimax raw

Графік навантажень



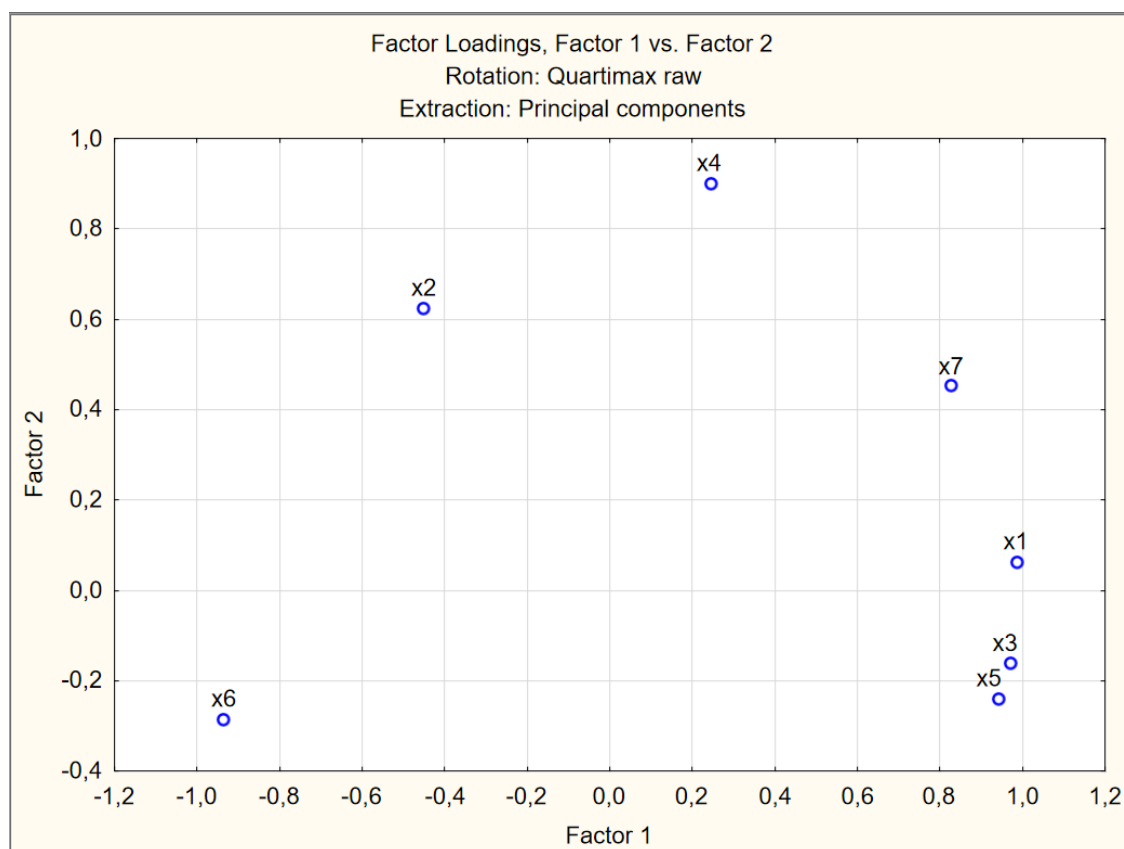
Навантаження факторів

Factor Loadings (Varimax raw) (Spreadsheet7)			
Extraction: Principal components			
(Marked loadings are >,700000)			
Variable	Factor 1	Factor 2	
x1	0,975956	-0,168932	
x2	-0,517459	-0,570495	
x3	0,983637	0,055233	
x4	0,147473	-0,921475	
x5	0,964989	0,135914	
x6	-0,898038	0,386967	
x7	0,772468	-0,539814	
Expl. Var	4,543926	1,665785	
Prp. Totl	0,649132	0,237969	

Повернення простору доволі велике.

Quartimax raw

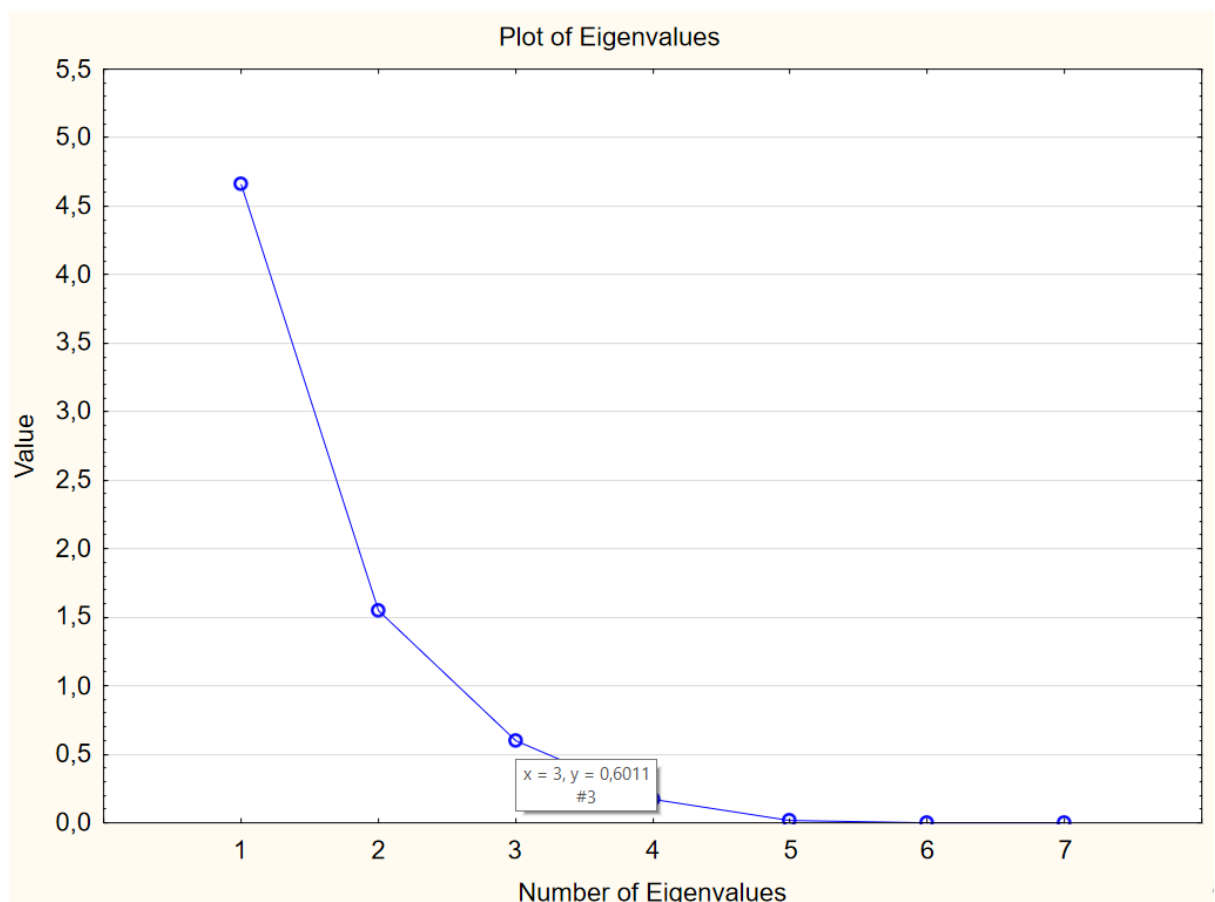
Графік навантажень



Навантаження факторів

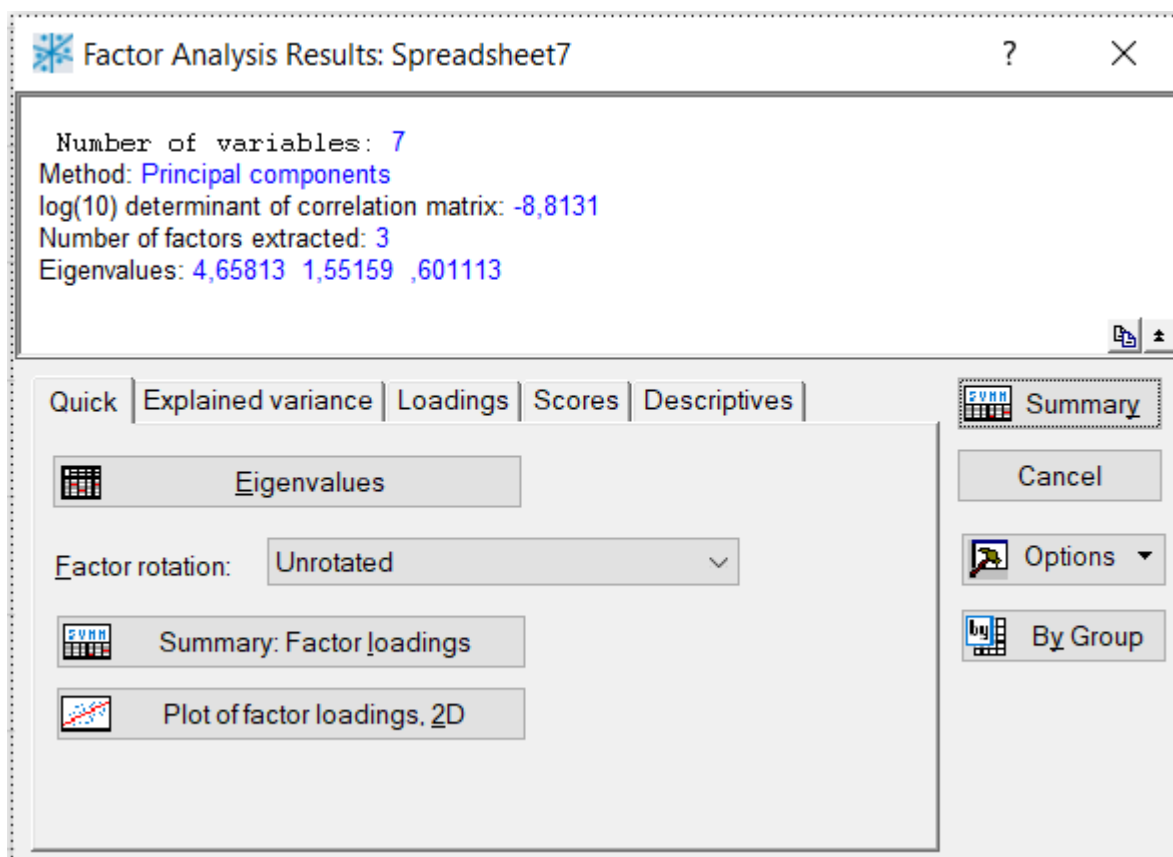
Factor Loadings (Quartimax raw) (Spreadsheet7) Extraction: Principal components (Marked loadings are >,700000)				
Variable	Factor 1	Factor 2		
x1	0,988611	0,060624		
x2	-0,451610	0,623921		
x3	0,971605	-0,163026		
x4	0,247874	0,899679		
x5	0,944200	-0,241169		
x6	-0,935134	-0,285903		
x7	0,827127	0,451628		
Expl.Var	4,636888	1,572823		
Prp.Totl	0,662413	0,224689		

Зовсім трохи зсунулася відносно не оберненого варіанту.



Згідно критерію Кеттела можемо залишити 2 або 3 змінні, бо з 4-ої в нас залишається майже пряма лінія. Хоча спад між 3 і 4 доволі малий. Тому, як на мене краще залишити 2 фактори.

Для значення 0.75 для критерію Кайзера нічого не повинно змінитись, ми так само будемо мати 2 критерія. Виставлю значення 0.6, щоб мати 3 критерія.

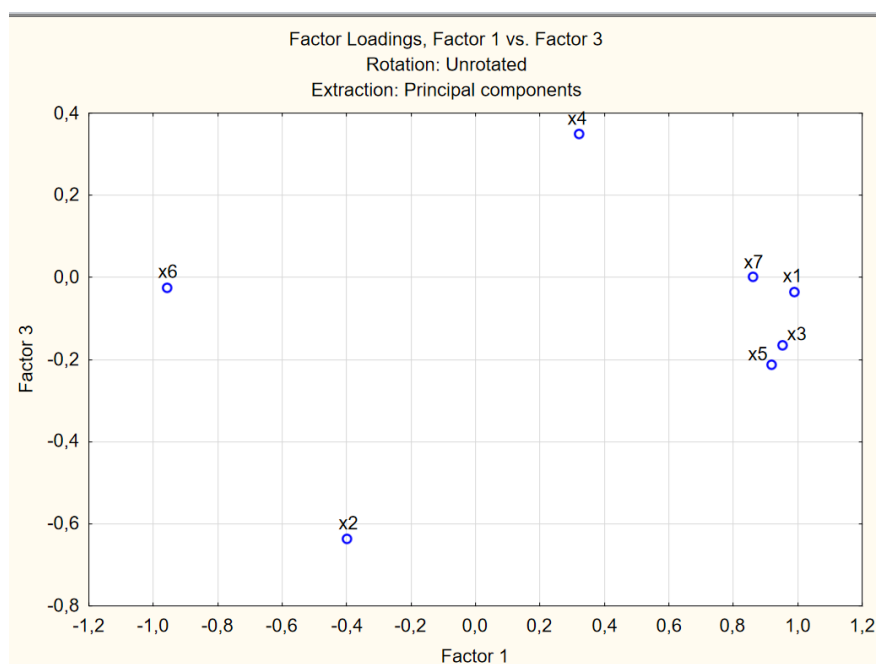
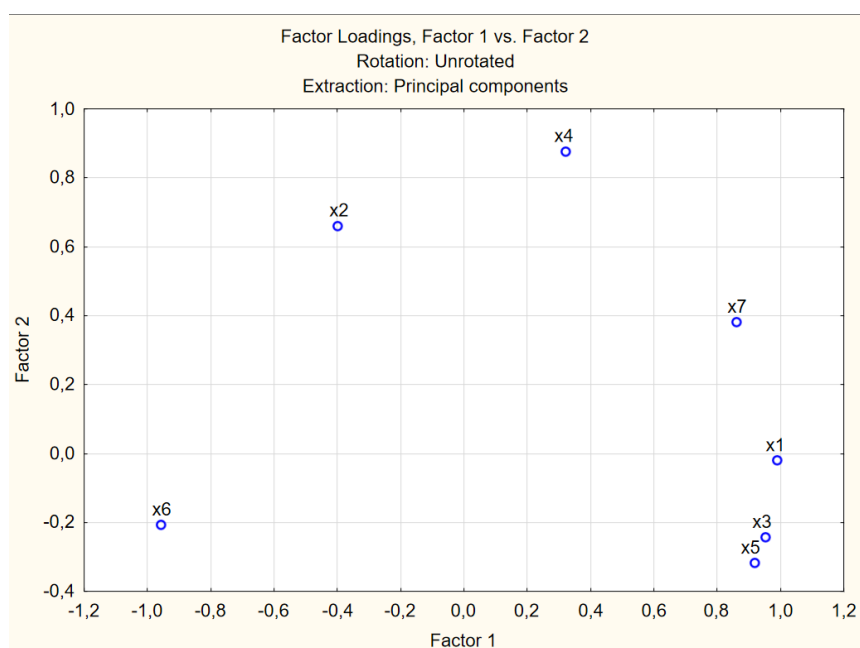


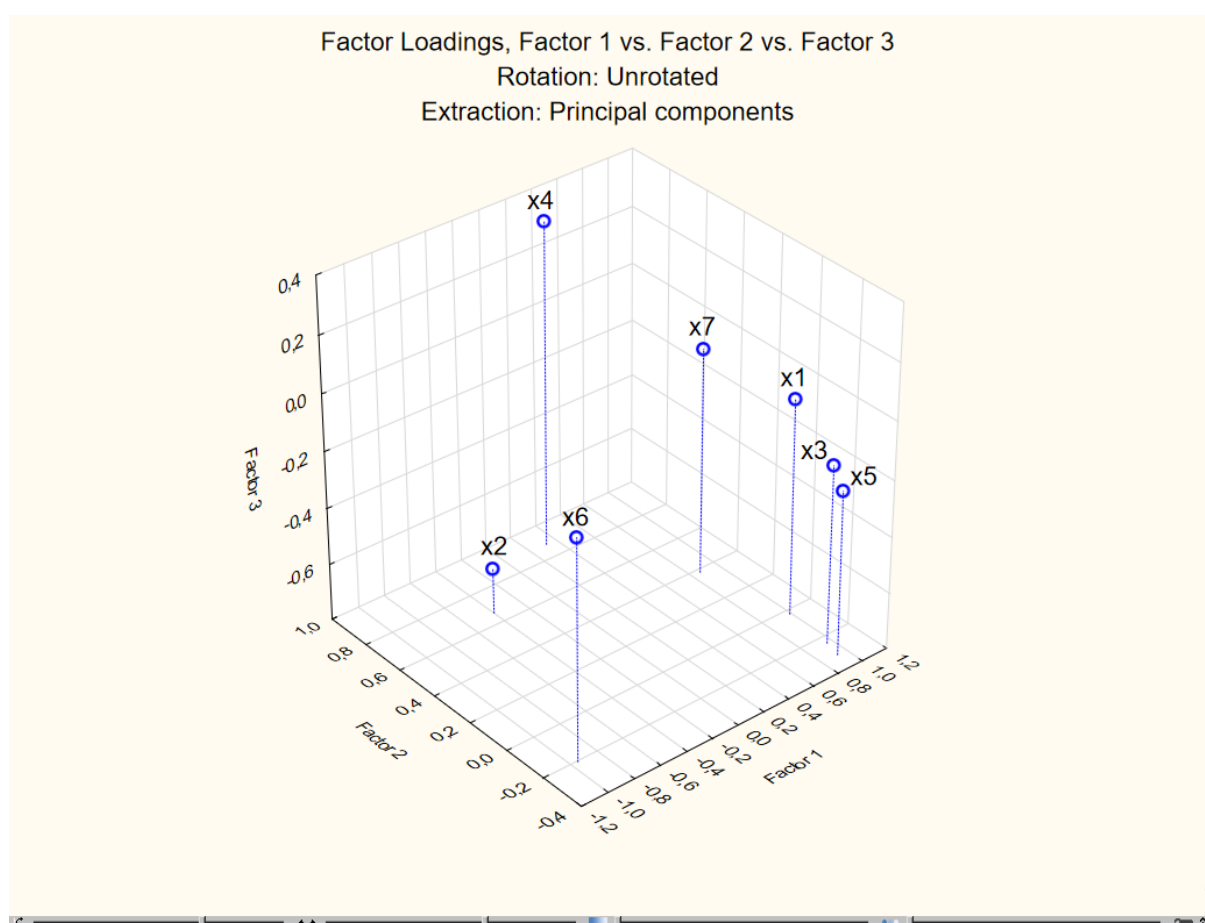
Таблиця факторних навантажень

Factor Loadings (Unrotated) (Spreadsheet7) Extraction: Principal components (Marked loadings are >,700000)				
Variable	Factor 1	Factor 2	Factor 3	
x1	0,990239	-0,021323	-0,035967	
x2	-0,398477	0,659124	-0,636334	
x3	0,954799	-0,242802	-0,165659	
x4	0,321412	0,876104	0,348017	
x5	0,921027	-0,318411	-0,213670	
x6	-0,955571	-0,207606	-0,026180	
x7	0,861636	0,381693	0,000167	
Expl. Var	4,658126	1,551586	0,601113	
Prp. Totl	0,665447	0,221655	0,085873	

Таблица спільностей

Communalities (Spreadsheet7)				
Extraction: Principal components				
Rotation: Unrotated				
Variable	From 1 Factor	From 2 Factors	From 3 Factors	Multiple R-Square
x1	0,980573	0,981027	0,982321	0,999368
x2	0,158784	0,593228	0,998149	0,434793
x3	0,911640	0,970593	0,998036	0,999922
x4	0,103306	0,870864	0,991980	0,948666
x5	0,848290	0,949676	0,995331	0,999854
x6	0,913116	0,956216	0,956902	0,990912
x7	0,742417	0,888106	0,888106	0,917200





За результатом додавання третього фактору в нас вийшло, що він не пояснює жодну змінну. Усі змінні здатні пояснити її 2 фактори.

Що ми бачимо в результаті вибуту критерія Кайзера на рівні 1. У нас є два фактори, перший пояснює змінні x1, x3, x5, x6, x7, другий x4. Схоже передбачення можна було зробити на етапі кореляцій. Але змінну x2, з якоїсь причини не може пояснити навіть 3-ій фактор. Але для 2-ох факторів другий пояснює її значно краще ніж перший(що також можна побачити на етапі кореляційної матриці).

Висновки:

В процесі виконання комп'ютерного практикуму №6 я виконав розрахунки згідно індивідуального завдання і набув вмінь факторного аналізу в Statistica. Було зручно отримувати результати кореляцій і факторного аналізу. Дуже просто виконувати обертання факторів різними методами ортогонального обертання. Легко отримати графіки і таблиці навантажень. В процесі роботи Statistica мала всі зручні показники.(можливо можна було б розфарбовувати кореляції автоматично в залежності від рівня кореляції)