

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 № 3
Кореляційно-регресійний аналіз у середовищі STATISTICA
з кредитного модуля «Багатовимірний статистичний аналіз»

Звіт склав
студент гр. КІ01
Копцов В.О.
Прийняла: *Ірина Джигирей*

Мета роботи. Розглянути види мір подібності. Опанувати методи розрахунку деяких мір відстані, використовуваних, зокрема, у кластерному аналізі. Набути вміння визначення таких мір відстані: евклідової, квадратичної евклідової, манхеттенівської та степеневої. Набути умінь використання базових алгоритмів кластеризування, набути вмінь проведення кластерного аналізу сукупності об'єктів за допомогою застосунку Statistica.

Завдання. За наданими даними спостережень виконати розрахунок евклідової, квадратичної евклідової, манхеттенівської та степеневої мір відстаней. Перевірити отримані результати за допомогою застосунку Statistica. Згідно індивідуального завдання (Додаток А) виконати кластеризування об'єктів за двома ознаками у застосунку Statistica.

Варіант №30

1.

	1	2	3
	x	y	z
1	60	90	30
2	90	90	60
3	150	30	90
4	150	60	120
5	180	60	90

Prüfung:

Berechnung:

$$d_{\text{unc}}(V_1, V_2, V_3) = \sqrt{(V_{1x} - V_{2x})^2 + (V_{1y} - V_{2y})^2 + (V_{1z} - V_{2z})^2}$$

$$d_{\text{unc}}(V_1, V_2) = \sqrt{30^2 + 0 + 30^2} = 30\sqrt{2}$$

$$d_{\text{unc}}(V_1, V_3) = \sqrt{90^2 + 60^2 + 60^2} = 30\sqrt{17}$$

$$d_{\text{unc}}(V_2, V_3) = \sqrt{60^2 + 60^2 + 30^2} = 90$$

$$d_{\text{unc}}(V_1, V_4) = \sqrt{90^2 + 30^2 + 90^2} = 30\sqrt{19}$$

$$d_{\text{unc}}(V_2, V_4) = \sqrt{60^2 + 30^2 + 60^2} = 90$$

$$d_{\text{unc}}(V_1, V_5) = \sqrt{90^2 + 90^2 + 50^2} = 30\sqrt{17}$$

$$d_{\text{unc}}(V_2, V_5) = \sqrt{30^2 + 0 + 30^2} = 30\sqrt{2}$$

$$d_{\text{unc}}(V_3, V_4) = \sqrt{0 + 30^2 + 30^2} = 30\sqrt{2}$$

$$d_{\text{unc}}(V_3, V_5) = \sqrt{720^2 + 30^2 + 60^2} = 30\sqrt{17}$$

$$d_{\text{unc}}(V_4, V_5) = \sqrt{30^2 + 30^2 + 0} = 30\sqrt{2}$$

Nearest neighbor

V_1	V_2	V_3	V_4	V_5
0	1800	0	0	1800
1800	0	8100	1800	1800
15300	8100	0	1800	1800
17100	8100	1800	0	1800
18900	9900	1800	1800	0

Handwritten notes:

$d_{\text{unc}}(V_1, V_2) = \sqrt{(V_{1x} - V_{2x})^2 + (V_{1y} - V_{2y})^2 + (V_{1z} - V_{2z})^2}$

$d_{\text{unc}}(V_1, V_2) = \sqrt{30^2 + 0 + 30^2} = 30\sqrt{2}$

$d_{\text{unc}}(V_1, V_3) = \sqrt{90^2 + 60^2 + 60^2} = 30\sqrt{17}$

$d_{\text{unc}}(V_2, V_3) = \sqrt{60^2 + 60^2 + 30^2} = 90$

$d_{\text{unc}}(V_1, V_4) = \sqrt{90^2 + 30^2 + 90^2} = 30\sqrt{19}$

$d_{\text{unc}}(V_2, V_4) = \sqrt{60^2 + 30^2 + 60^2} = 90$

$d_{\text{unc}}(V_1, V_5) = \sqrt{90^2 + 90^2 + 50^2} = 30\sqrt{17}$

$d_{\text{unc}}(V_2, V_5) = \sqrt{30^2 + 0 + 30^2} = 30\sqrt{2}$

$d_{\text{unc}}(V_3, V_4) = \sqrt{0 + 30^2 + 30^2} = 30\sqrt{2}$

$d_{\text{unc}}(V_3, V_5) = \sqrt{720^2 + 30^2 + 60^2} = 30\sqrt{17}$

$d_{\text{unc}}(V_4, V_5) = \sqrt{30^2 + 30^2 + 0} = 30\sqrt{2}$

Handwritten notes:

$d_{\text{unc}}(V_1, V_2) = \sqrt{(V_{1x} - V_{2x})^2 + (V_{1y} - V_{2y})^2 + (V_{1z} - V_{2z})^2}$

$d_{\text{unc}}(V_1, V_2) = \sqrt{30^2 + 0 + 30^2} = 30\sqrt{2}$

$d_{\text{unc}}(V_1, V_3) = \sqrt{90^2 + 60^2 + 60^2} = 30\sqrt{17}$

$d_{\text{unc}}(V_2, V_3) = \sqrt{60^2 + 60^2 + 30^2} = 90$

$d_{\text{unc}}(V_1, V_4) = \sqrt{90^2 + 30^2 + 90^2} = 30\sqrt{19}$

$d_{\text{unc}}(V_2, V_4) = \sqrt{60^2 + 30^2 + 60^2} = 90$

$d_{\text{unc}}(V_1, V_5) = \sqrt{90^2 + 90^2 + 50^2} = 30\sqrt{17}$

$d_{\text{unc}}(V_2, V_5) = \sqrt{30^2 + 0 + 30^2} = 30\sqrt{2}$

$d_{\text{unc}}(V_3, V_4) = \sqrt{0 + 30^2 + 30^2} = 30\sqrt{2}$

$d_{\text{unc}}(V_3, V_5) = \sqrt{720^2 + 30^2 + 60^2} = 30\sqrt{17}$

$d_{\text{unc}}(V_4, V_5) = \sqrt{30^2 + 30^2 + 0} = 30\sqrt{2}$

$d_{c-b}(V_i, V_j) = |V_{ix} - V_{jx}| + |V_{iy} - V_{jy}| + |V_{iz} - V_{jz}|$

$d_{c-b}(V_1, V_2) = 30 + 0 + 30 = 60$
 $d_{c-b}(V_1, V_3) = 90 + 60 + 60 = 210$
 $d_{c-b}(V_2, V_3) = 60 + 60 + 30 = 150$
 $d_{c-b}(V_1, V_4) = 90 + 30 + 90 = 210$
 $d_{c-b}(V_2, V_4) = 60 + 30 + 60 = 150$
 $d_{c-b}(V_3, V_4) = 0 + 30 + 30 = 60$
 $d_{c-b}(V_2, V_5) = 90 + 30 + 30 = 150$
 $d_{c-b}(V_4, V_5) = 30 + 30 + 0 = 60$

$(p=2, r=3)$ $d_{pow}(V_i, V_j) = \sqrt[p]{|V_{ix} - V_{jx}|^p + |V_{iy} - V_{jy}|^p + |V_{iz} - V_{jz}|^p}$
 $d_{pow} = \sqrt[p]{(d_{c-b})^p}$

$d_{pow}(V_1, V_2) = \sqrt[2]{60^2} = 60$
 $d_{pow}(V_1, V_3) = \sqrt[2]{210^2} = 210$
 $d_{pow}(V_2, V_3) = \sqrt[2]{150^2} = 150$
 $d_{pow}(V_1, V_4) = \sqrt[2]{210^2} = 210$
 $d_{pow}(V_2, V_4) = \sqrt[2]{150^2} = 150$
 $d_{pow}(V_3, V_4) = \sqrt[2]{60^2} = 60$
 $d_{pow}(V_2, V_5) = \sqrt[2]{150^2} = 150$
 $d_{pow}(V_4, V_5) = \sqrt[2]{60^2} = 60$

2.

Case No.	Euclidean distances (Spreadsheet7)				
	C_1	C_2	C_3	C_4	C_5
C_1	0	42,4	124	131	137
C_2	42	0,0	90	90	99
C_3	124	90,0	0	42	42
C_4	131	90,0	42	0	42
C_5	137	99,5	42	42	0

Case No.	Squared Euclidean distances (Spreadsheet7)				
	C_1	C_2	C_3	C_4	C_5
C_1	0	1800	15300	17100	18900
C_2	1800	0	8100	8100	9900
C_3	15300	8100	0	1800	1800
C_4	17100	8100	1800	0	1800
C_5	18900	9900	1800	1800	0

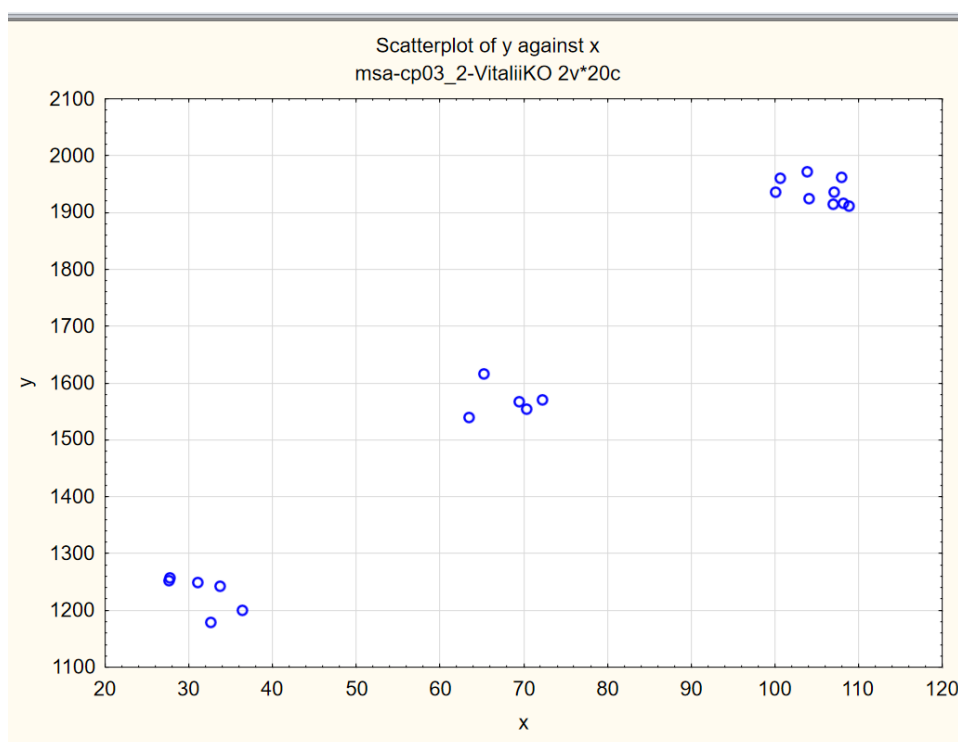
Case No.	City-block (Manhattan) distances (Spreadsheet7)				
	C_1	C_2	C_3	C_4	C_5
C_1	0	60	210	210	210
C_2	60	0	150	150	150
C_3	210	150	0	60	60
C_4	210	150	60	0	60
C_5	210	150	60	60	0

Case No.	Power: SUM(ABS(x-y)**p)**1/r (Spreadsheet7)				
	C_1	C_2	C_3	C_4	C_5
C_1	0,0	12,2	24,8	25,8	26,6
C_2	12,2	0,0	20,1	20,1	21,5
C_3	24,8	20,1	0,0	12,2	12,2
C_4	25,8	20,1	12,2	0,0	12,2
C_5	26,6	21,5	12,2	12,2	0,0

3. В результаті матриці розраховані вручну для манхеттенівської і квадратичної евклідової відстані повністю співпадають з матрицями розрахованими в статистиці, для інших видів відстаней результати також співпадають(але в статистиці вони округлені). В результаті зрозуміло - розраховувати таке за допомогою статистики набагато легше, ніж вручну.

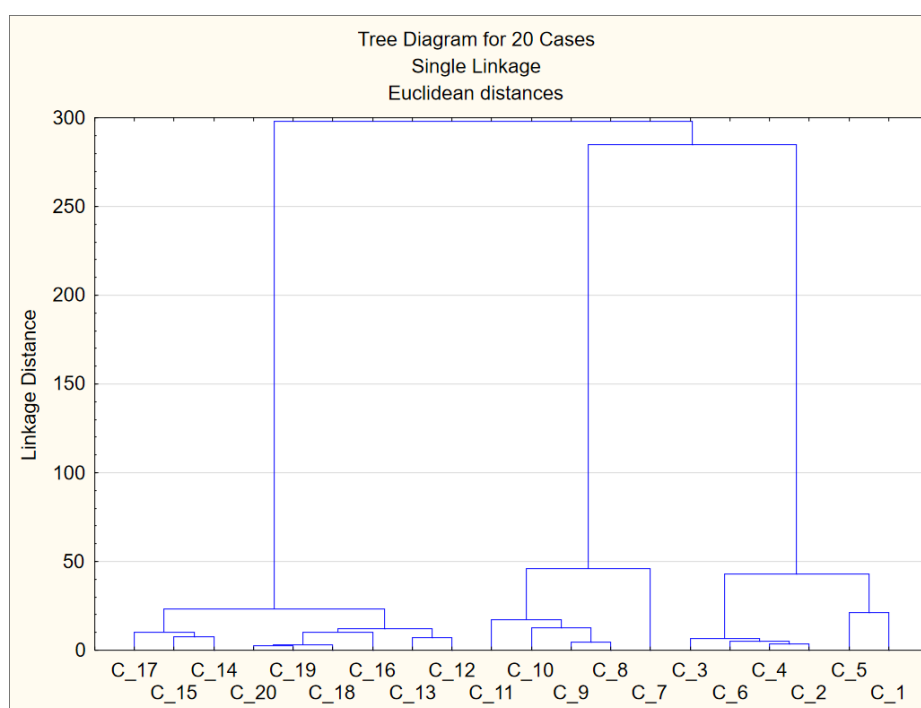
4. Метод близького сусіда(евклідова відстань)

Звичайні дані



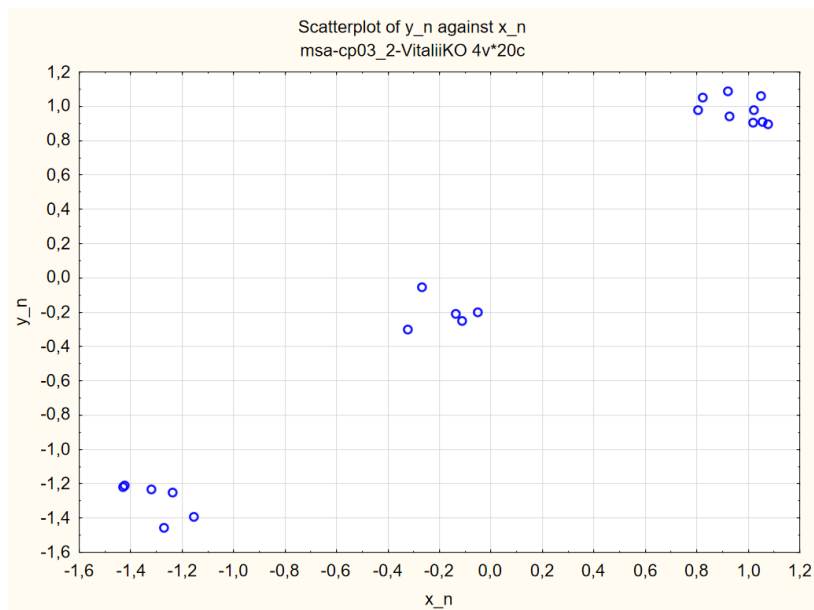
Case No.	Euclidean distances (msa-cp03_2-VitaliiKO)													
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_13	C_14
C_1	0,00	73,99	63,92	77,57	21,22	70,00	438,40	393,45	389,76	377,49	361,11	760,55	761,23	783,68
C_2	73,99	0,00	11,69	3,60	53,67	5,22	365,32	320,77	316,99	304,81	288,23	687,58	688,38	710,69
C_3	63,92	11,69	0,00	14,80	43,11	6,63	374,60	329,80	326,06	313,83	297,37	696,81	697,54	719,93
C_4	77,57	3,60	14,80	0,00	57,20	8,17	361,73	317,19	313,40	301,23	284,64	683,99	684,79	707,09
C_5	21,22	53,67	43,11	57,20	0,00	49,39	417,30	372,30	368,62	356,34	339,99	739,42	740,08	762,56
C_6	70,00	5,22	6,63	8,17	49,39	0,00	368,79	324,10	320,34	308,13	291,62	691,03	691,79	714,14
C_7	438,40	365,32	374,60	361,73	417,30	368,79	0,00	46,27	49,37	61,79	77,41	322,26	323,12	345,36
C_8	393,45	320,77	329,80	317,19	372,30	324,10	46,27	0,00	4,39	15,96	32,86	367,16	367,78	390,31
C_9	389,76	316,99	326,06	313,40	368,62	320,34	49,37	4,39	0,00	12,43	28,86	370,81	371,48	393,95
C_10	377,49	304,81	313,83	301,23	356,34	308,13	61,79	15,96	12,43	0,00	17,24	383,11	383,74	406,25
C_11	361,11	288,23	297,37	284,64	339,99	291,62	77,41	32,86	28,86	17,24	0,00	399,44	400,17	422,57
C_12	760,55	687,58	696,81	683,99	739,42	691,03	322,26	367,16	370,81	383,11	399,44	0,00	7,04	23,18
C_13	761,23	688,38	697,54	684,79	740,08	691,79	323,12	367,78	371,48	383,74	400,17	7,04	0,00	24,04
C_14	783,68	710,69	719,93	707,09	762,56	714,14	345,36	390,31	393,95	406,25	422,57	23,18	24,04	0,00
C_15	786,67	713,79	722,96	710,20	765,53	717,21	348,50	393,23	396,92	409,18	425,59	26,69	25,49	7,68
C_16	749,35	676,47	685,64	672,87	728,21	679,88	311,18	355,91	359,59	371,86	388,27	12,28	12,04	34,96
C_17	795,64	722,68	731,90	719,08	774,51	726,12	357,36	402,24	405,90	418,19	434,53	35,09	35,02	12,15
C_18	736,14	663,38	672,48	659,79	714,99	666,76	298,21	342,69	346,42	358,66	375,15	26,85	25,45	49,23
C_19	740,66	667,88	676,99	664,28	719,51	671,26	302,68	347,21	350,93	363,17	379,65	22,30	20,81	44,58
C_20	738,51	665,71	674,83	662,11	717,36	669,10	300,49	345,06	348,77	361,02	377,48	23,83	22,82	46,41

Отримали матрицю відстаней



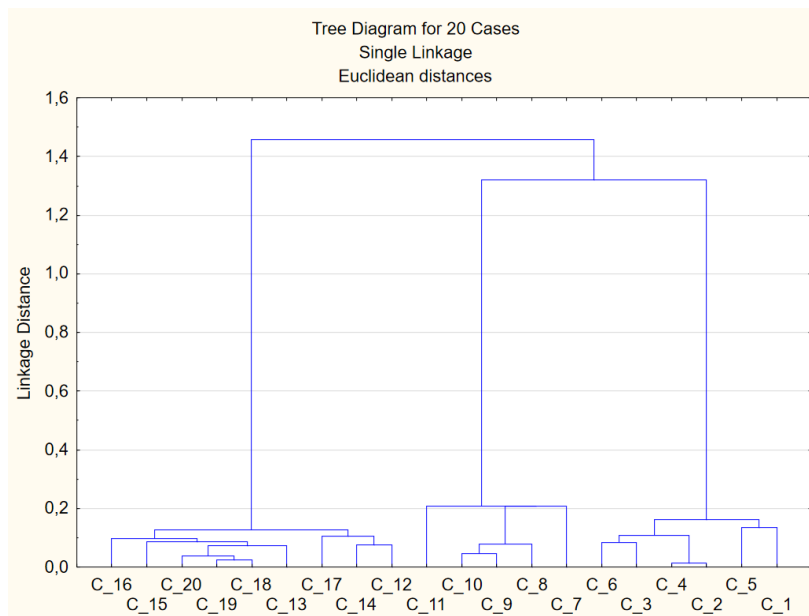
Вертикальна дендрограма

Нормалізовані дані



Case No.	Euclidean distances (msa-cp03_2-VitaliiKO)																
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_13	C_14	C_15	C_16	C_17
C_1	0.00	0.28	0.21	0.29	0.13	0.23	1.73	1.75	1.69	1.67	1.50	3.20	3.35	3.27	3.42	3.25	3.36
C_2	0.28	0.00	0.19	0.01	0.32	0.11	1.65	1.71	1.64	1.64	1.44	3.13	3.29	3.20	3.37	3.20	3.30
C_3	0.21	0.19	0.00	0.19	0.16	0.08	1.54	1.58	1.52	1.51	1.32	3.02	3.17	3.09	3.25	3.08	3.19
C_4	0.29	0.01	0.19	0.00	0.32	0.11	1.63	1.70	1.63	1.63	1.43	3.12	3.28	3.19	3.36	3.19	3.28
C_5	0.13	0.32	0.16	0.32	0.00	0.23	1.61	1.62	1.56	1.55	1.37	3.07	3.22	3.14	3.30	3.13	3.24
C_6	0.23	0.11	0.08	0.11	0.23	0.00	1.58	1.63	1.56	1.56	1.36	3.07	3.22	3.13	3.30	3.13	3.23
C_7	1.73	1.65	1.54	1.63	1.61	1.58	0.00	0.26	0.21	0.25	0.25	1.49	1.65	1.55	1.72	1.56	1.65
C_8	1.75	1.71	1.58	1.70	1.62	1.63	0.26	0.00	0.08	0.08	0.29	1.46	1.59	1.53	1.67	1.50	1.62
C_9	1.69	1.64	1.52	1.63	1.56	1.56	0.21	0.08	0.00	0.05	0.21	1.52	1.66	1.59	1.74	1.57	1.68
C_10	1.67	1.64	1.51	1.63	1.55	1.56	0.25	0.08	0.05	0.00	0.22	1.53	1.67	1.60	1.75	1.58	1.69
C_11	1.50	1.44	1.32	1.43	1.37	1.36	0.25	0.29	0.21	0.22	0.00	1.71	1.86	1.77	1.93	1.76	1.87
C_12	3.20	3.13	3.02	3.12	3.07	3.07	1.49	1.46	1.52	1.53	1.71	0.00	0.22	0.08	0.26	0.13	0.16
C_13	3.35	3.29	3.17	3.28	3.22	3.22	1.65	1.59	1.66	1.67	1.86	0.22	0.00	0.21	0.09	0.10	0.15
C_14	3.27	3.20	3.09	3.19	3.14	3.13	1.55	1.53	1.59	1.60	1.77	0.08	0.21	0.00	0.23	0.15	0.11
C_15	3.42	3.37	3.25	3.36	3.30	3.30	1.72	1.67	1.74	1.75	1.93	0.26	0.09	0.23	0.00	0.17	0.13
C_16	3.25	3.20	3.08	3.19	3.13	3.13	1.56	1.50	1.57	1.58	1.76	0.13	0.10	0.15	0.17	0.00	0.15
C_17	3.36	3.30	3.19	3.28	3.24	3.23	1.65	1.62	1.68	1.69	1.87	0.16	0.15	0.11	0.13	0.15	0.00
C_18	3.32	3.28	3.16	3.27	3.20	3.21	1.65	1.57	1.64	1.65	1.84	0.28	0.10	0.30	0.17	0.15	0.25
C_19	3.32	3.27	3.15	3.26	3.19	3.20	1.64	1.57	1.64	1.65	1.84	0.26	0.07	0.27	0.15	0.13	0.22
C_20	3.29	3.24	3.12	3.23	3.16	3.17	1.60	1.54	1.61	1.62	1.80	0.23	0.07	0.25	0.16	0.10	0.21

Матриця відстаней

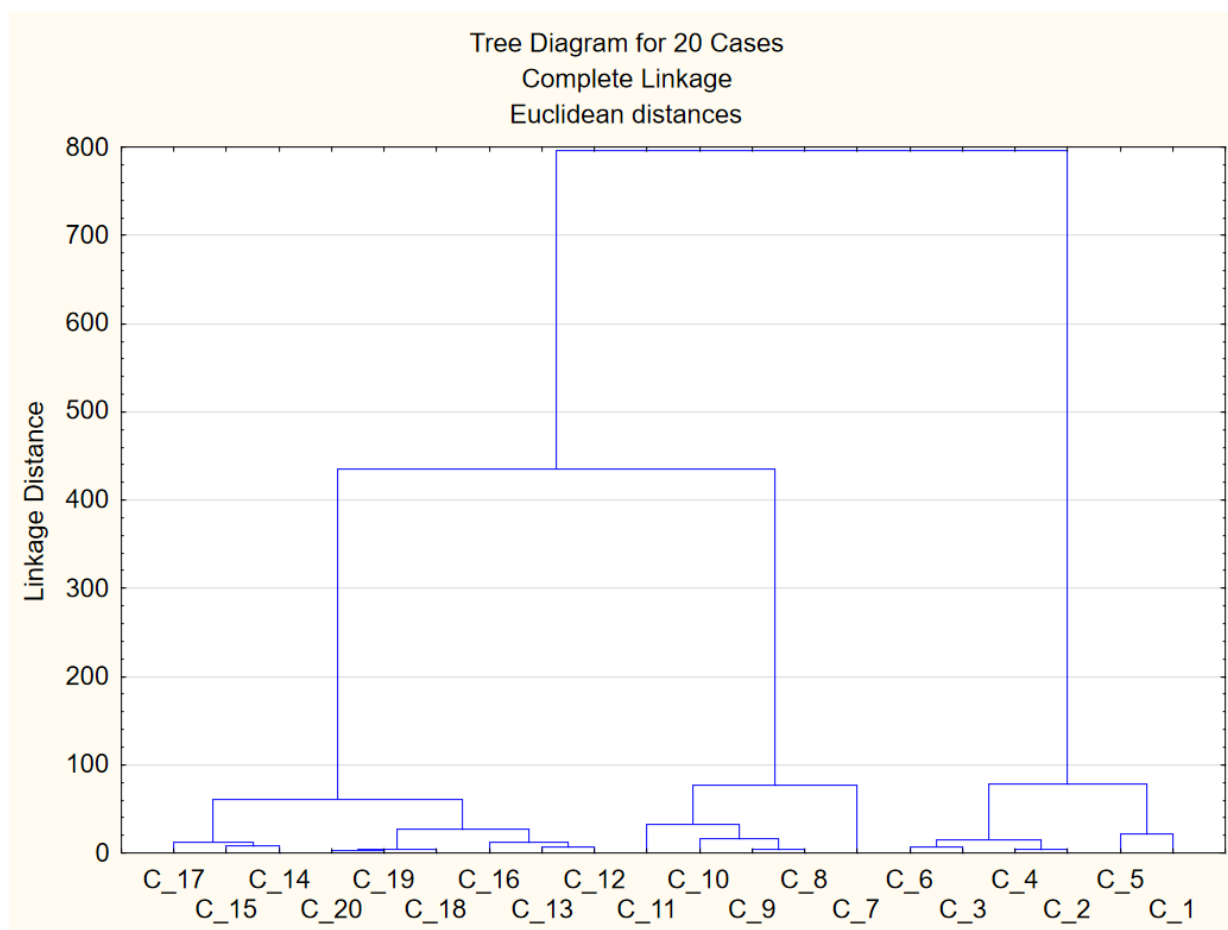


Вертикальна дендрограма

5. Метод далекого сусіда(евклідова відстань)

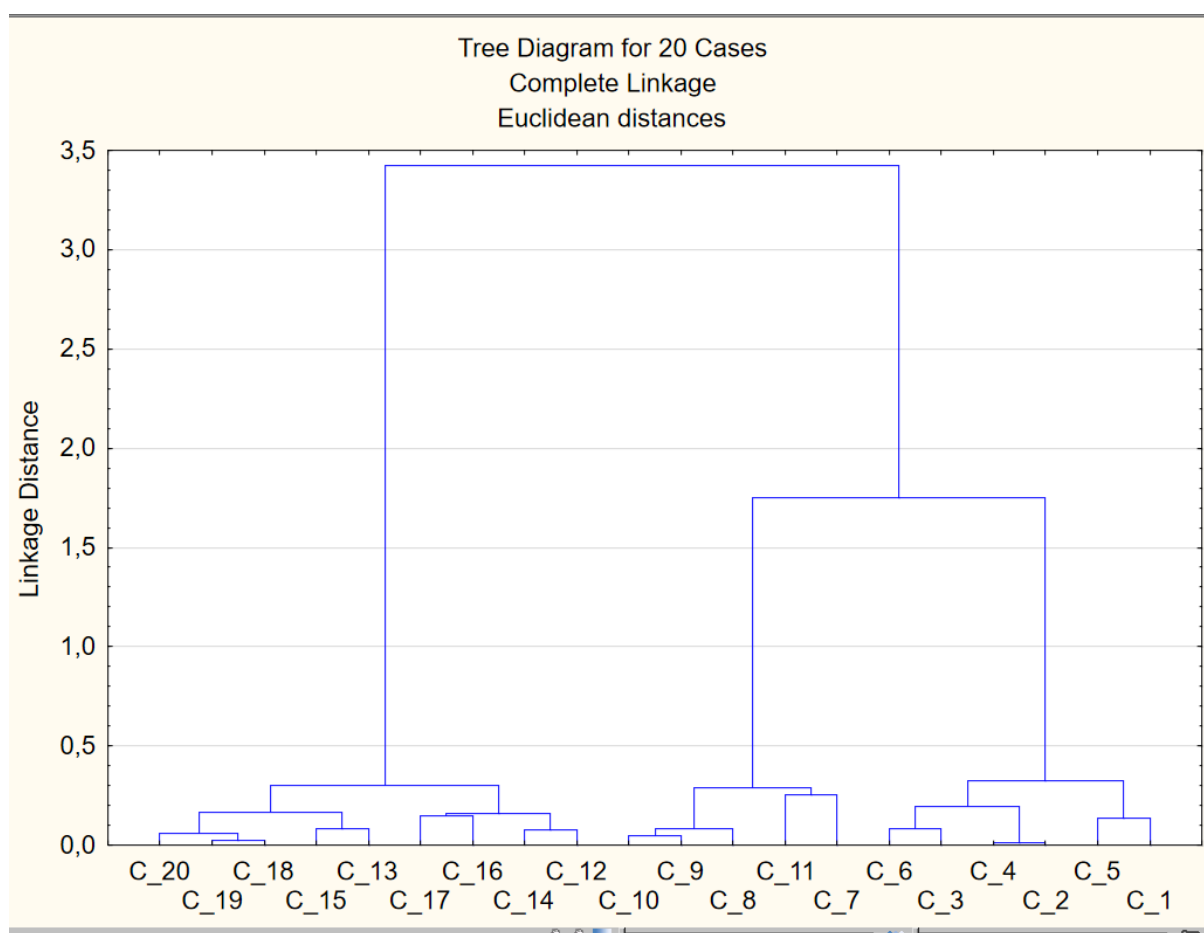
Нормалізовані дані

Case No.	Euclidean distances (msa-cp03_2-VitaliiKO)														
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_13	C_14	C_15
C_1	0,00	73,99	63,92	77,57	21,22	70,00	438,40	393,45	389,76	377,49	361,11	760,55	761,23	783,68	786,67
C_2	73,99	0,00	11,69	3,60	53,67	5,22	365,32	320,77	316,99	304,81	288,23	687,58	688,38	710,69	713,79
C_3	63,92	11,69	0,00	14,80	43,11	6,63	374,60	329,80	326,06	313,83	297,37	696,81	697,54	719,93	722,96
C_4	77,57	3,60	14,80	0,00	57,20	8,17	361,73	317,19	313,40	301,23	284,64	683,99	684,79	707,09	710,20
C_5	21,22	53,67	43,11	57,20	0,00	49,39	417,30	372,30	368,62	356,34	339,99	739,42	740,08	762,56	765,53
C_6	70,00	5,22	6,63	8,17	49,39	0,00	368,79	324,10	320,34	308,13	291,62	691,03	691,79	714,14	717,21
C_7	438,40	365,32	374,60	361,73	417,30	368,79	0,00	46,27	49,37	61,79	77,41	322,26	323,12	345,36	348,50
C_8	393,45	320,77	329,80	317,19	372,30	324,10	46,27	0,00	4,39	15,96	32,86	367,16	367,78	390,31	393,23
C_9	389,76	316,99	326,06	313,40	368,62	320,34	49,37	4,39	0,00	12,43	28,86	370,81	371,48	393,95	396,92
C_10	377,49	304,81	313,83	301,23	356,34	308,13	61,79	15,96	12,43	0,00	17,24	383,11	383,74	406,25	409,18
C_11	361,11	288,23	297,37	284,64	339,99	291,62	77,41	32,86	28,86	17,24	0,00	399,44	400,17	422,57	425,59
C_12	760,55	687,58	696,81	683,99	739,42	691,03	322,26	367,16	370,81	383,11	399,44	0,00	7,04	23,18	26,69
C_13	761,23	688,38	697,54	684,79	740,08	691,79	323,12	367,78	371,48	383,74	400,17	7,04	0,00	24,04	25,49
C_14	783,68	710,69	719,93	707,09	762,56	714,14	345,36	390,31	393,95	406,25	422,57	23,18	24,04	0,00	7,68
C_15	786,67	713,79	722,96	710,20	765,53	717,21	348,50	393,23	396,92	409,18	425,59	26,69	25,49	7,68	0,00
C_16	749,35	676,47	685,64	672,87	728,21	679,88	311,18	355,91	359,59	371,86	388,27	12,28	12,04	34,96	37,32
C_17	795,64	722,68	731,90	719,08	774,51	726,12	357,36	402,24	405,90	418,19	434,53	35,09	35,02	12,15	10,25
C_18	736,14	663,38	672,48	659,79	714,99	666,76	298,21	342,69	346,42	358,66	375,15	26,85	25,45	49,23	50,88
C_19	740,66	667,88	676,99	664,28	719,51	671,26	302,68	347,21	350,93	363,17	379,65	22,30	20,81	44,58	46,26
C_20	738,51	665,71	674,83	662,11	717,36	669,10	300,49	345,06	348,77	361,02	377,48	23,83	22,82	46,41	48,31



Нормалізовані дані

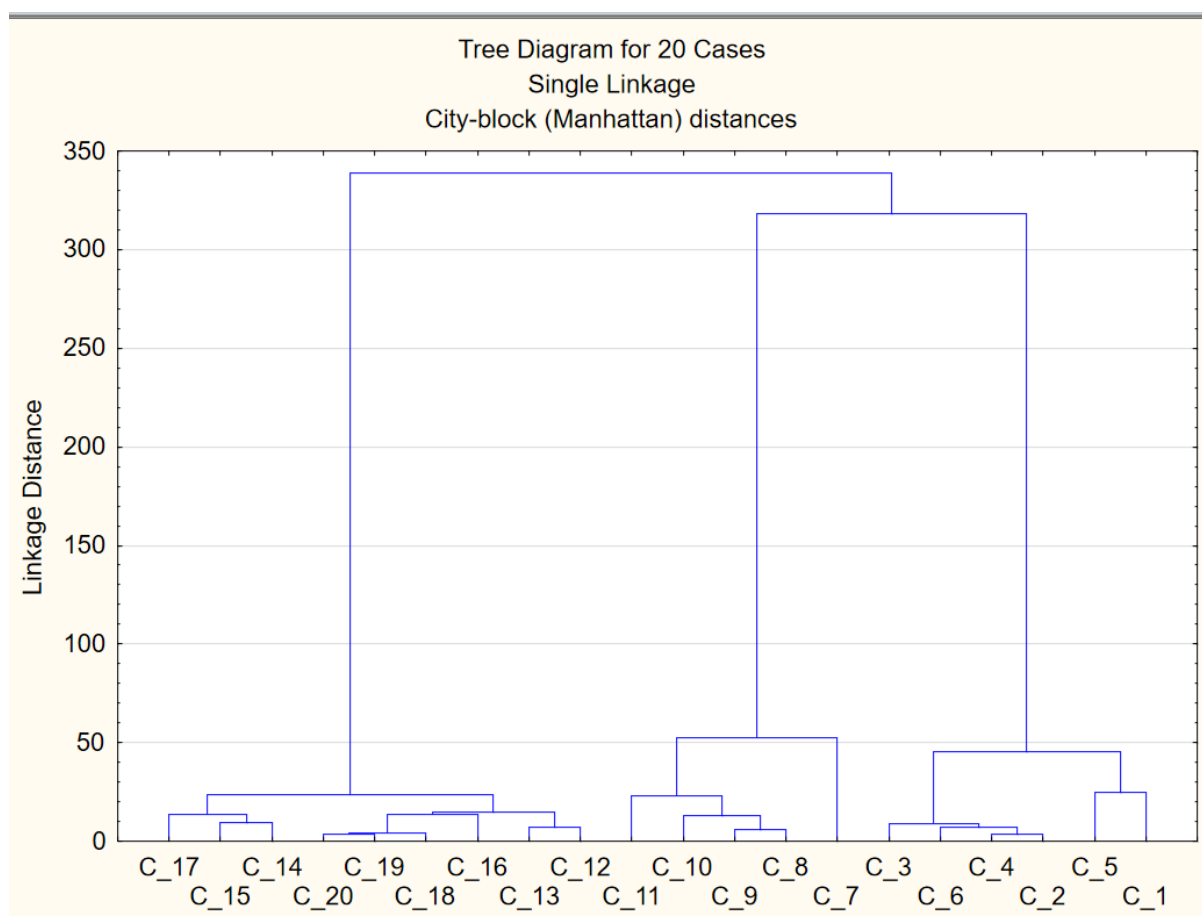
Case No.	Euclidean distances (msa-cp03_2-VitaliiKO)														
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_13	C_14	C_15
C_1	0,00	0,28	0,21	0,29	0,13	0,23	1,73	1,75	1,69	1,67	1,50	3,20	3,35	3,27	3,42
C_2	0,28	0,00	0,19	0,01	0,32	0,11	1,65	1,71	1,64	1,64	1,44	3,13	3,29	3,20	3,37
C_3	0,21	0,19	0,00	0,19	0,16	0,08	1,54	1,58	1,52	1,51	1,32	3,02	3,17	3,09	3,25
C_4	0,29	0,01	0,19	0,00	0,32	0,11	1,63	1,70	1,63	1,63	1,43	3,12	3,28	3,19	3,36
C_5	0,13	0,32	0,16	0,32	0,00	0,23	1,61	1,62	1,56	1,55	1,37	3,07	3,22	3,14	3,30
C_6	0,23	0,11	0,08	0,11	0,23	0,00	1,58	1,63	1,56	1,56	1,36	3,07	3,22	3,13	3,30
C_7	1,73	1,65	1,54	1,63	1,61	1,58	0,00	0,26	0,21	0,25	0,25	1,49	1,65	1,55	1,72
C_8	1,75	1,71	1,58	1,70	1,62	1,63	0,26	0,00	0,08	0,08	0,29	1,46	1,59	1,53	1,67
C_9	1,69	1,64	1,52	1,63	1,56	1,56	0,21	0,08	0,00	0,05	0,21	1,52	1,66	1,59	1,74
C_10	1,67	1,64	1,51	1,63	1,55	1,56	0,25	0,08	0,05	0,00	0,22	1,53	1,67	1,60	1,75
C_11	1,50	1,44	1,32	1,43	1,37	1,36	0,25	0,29	0,21	0,22	0,00	1,71	1,86	1,77	1,93
C_12	3,20	3,13	3,02	3,12	3,07	3,07	1,49	1,46	1,52	1,53	1,71	0,00	0,22	0,08	0,26
C_13	3,35	3,29	3,17	3,28	3,22	3,22	1,65	1,59	1,66	1,67	1,86	0,22	0,00	0,21	0,09
C_14	3,27	3,20	3,09	3,19	3,14	3,13	1,55	1,53	1,59	1,60	1,77	0,08	0,21	0,00	0,23
C_15	3,42	3,37	3,25	3,36	3,30	3,30	1,72	1,67	1,74	1,75	1,93	0,26	0,09	0,23	0,00
C_16	3,25	3,20	3,08	3,19	3,13	3,13	1,56	1,50	1,57	1,58	1,76	0,13	0,10	0,15	0,17
C_17	3,36	3,30	3,19	3,28	3,24	3,23	1,65	1,62	1,68	1,69	1,87	0,16	0,15	0,11	0,13
C_18	3,32	3,28	3,16	3,27	3,20	3,21	1,65	1,57	1,64	1,65	1,84	0,28	0,10	0,30	0,17
C_19	3,32	3,27	3,15	3,26	3,19	3,20	1,64	1,57	1,64	1,65	1,84	0,26	0,07	0,27	0,15
C_20	3,29	3,24	3,12	3,23	3,16	3,17	1,60	1,54	1,61	1,62	1,80	0,23	0,07	0,25	0,16



6. Метод близького сусіда(манхеттенівська відстань)

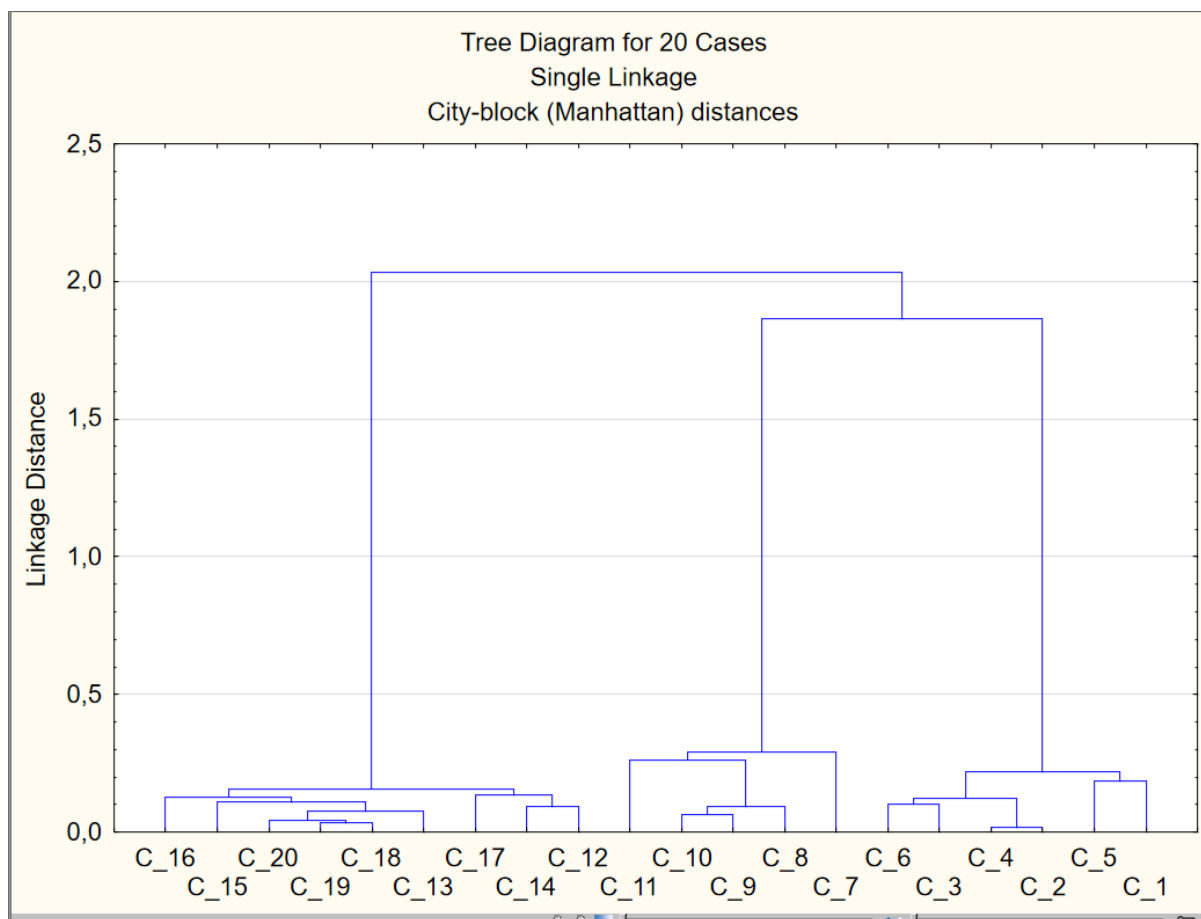
Нормалізовані дані

Case No.	City-block (Manhattan) distances (msa-cp03_2-VitaliiKO)																			
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_13	C_14	C_15	C_16	C_17	C_18	C_19	C_20
C_1	0	79	65	82	25	72	470	431	425	413	391	825	832	849	858	817	864	808	812	809
C_2	79	0	16	4	62	7	401	362	356	345	322	756	763	780	790	749	795	740	744	740
C_3	65	16	0	20	46	9	405	366	360	348	326	760	767	784	793	752	799	743	747	744
C_4	82	4	20	0	65	11	397	359	352	341	318	753	760	776	786	745	791	736	740	737
C_5	25	62	46	65	0	54	445	406	400	389	366	800	807	824	834	793	839	784	788	784
C_6	72	7	9	11	54	0	401	363	356	345	322	757	764	780	790	749	795	740	744	741
C_7	470	401	405	397	445	401	0	53	53	67	79	355	362	379	389	348	394	339	343	339
C_8	431	362	366	359	406	363	53	0	6	18	40	394	401	418	427	386	433	377	381	378
C_9	425	356	360	352	400	356	53	6	0	13	34	400	407	424	433	392	439	384	387	384
C_10	413	345	348	341	389	345	67	18	13	0	23	412	419	435	445	404	450	395	399	396
C_11	391	322	326	318	366	322	79	40	34	23	0	434	441	458	468	427	473	418	422	419
C_12	825	756	760	753	800	757	355	394	400	412	434	0	7	24	33	16	39	34	29	30
C_13	832	763	767	760	807	764	362	401	407	419	441	7	0	30	26	15	38	27	22	23
C_14	849	780	784	776	824	780	379	418	424	435	458	24	30	0	10	38	15	57	52	52
C_15	858	790	793	786	834	790	389	427	433	445	468	33	26	10	0	41	13	52	47	49
C_16	817	749	752	745	793	749	348	386	392	404	427	16	15	38	41	0	47	19	13	14
C_17	864	795	799	791	839	795	394	433	439	450	473	39	38	15	13	47	0	65	60	61
C_18	808	740	743	736	784	740	339	377	384	395	418	34	27	57	52	19	65	0	5	4
C_19	812	744	747	740	788	744	343	381	387	399	422	29	22	52	47	13	60	5	0	3
C_20	809	740	744	737	784	741	339	378	384	396	419	30	23	52	49	14	61	4	3	0



Нормалізовані дані

Case No.	City-block (Manhattan) distances (msa-cp03_2-VitaliiKO)															
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_13	C_14	C_15	C_16
C_1	0,00	0,39	0,24	0,40	0,18	0,27	2,41	2,48	2,38	2,37	2,10	4,51	4,73	4,60	4,84	4,60
C_2	0,39	0,00	0,22	0,02	0,44	0,12	2,33	2,40	2,30	2,29	2,02	4,43	4,65	4,52	4,76	4,52
C_3	0,24	0,22	0,00	0,23	0,22	0,10	2,17	2,24	2,14	2,13	1,86	4,27	4,49	4,36	4,60	4,36
C_4	0,40	0,02	0,23	0,00	0,45	0,13	2,31	2,38	2,29	2,27	2,01	4,42	4,63	4,51	4,74	4,50
C_5	0,18	0,44	0,22	0,45	0,00	0,32	2,22	2,29	2,20	2,18	1,92	4,33	4,55	4,42	4,65	4,41
C_6	0,27	0,12	0,10	0,13	0,32	0,00	2,23	2,30	2,21	2,19	1,93	4,34	4,55	4,43	4,66	4,42
C_7	2,41	2,33	2,17	2,31	2,22	2,23	0,00	0,36	0,29	0,36	0,30	2,10	2,32	2,20	2,43	2,19
C_8	2,48	2,40	2,24	2,38	2,29	2,30	0,36	0,00	0,09	0,11	0,37	2,04	2,25	2,13	2,36	2,12
C_9	2,38	2,30	2,14	2,29	2,20	2,21	0,29	0,09	0,00	0,06	0,28	2,13	2,35	2,22	2,45	2,22
C_10	2,37	2,29	2,13	2,27	2,18	2,19	0,36	0,11	0,06	0,00	0,26	2,14	2,36	2,24	2,47	2,23
C_11	2,10	2,02	1,86	2,01	1,92	1,93	0,30	0,37	0,28	0,26	0,00	2,41	2,62	2,50	2,73	2,49
C_12	4,51	4,43	4,27	4,42	4,33	4,34	2,10	2,04	2,13	2,14	2,41	0,00	0,22	0,09	0,32	0,16
C_13	4,73	4,65	4,49	4,63	4,55	4,55	2,32	2,25	2,35	2,36	2,62	0,22	0,00	0,27	0,11	0,13
C_14	4,60	4,52	4,36	4,51	4,42	4,43	2,20	2,13	2,22	2,24	2,50	0,09	0,27	0,00	0,23	0,22
C_15	4,84	4,76	4,60	4,74	4,65	4,66	2,43	2,36	2,45	2,47	2,73	0,32	0,11	0,23	0,00	0,24
C_16	4,60	4,52	4,36	4,50	4,41	4,42	2,19	2,12	2,22	2,23	2,49	0,16	0,13	0,22	0,24	0,00
C_17	4,74	4,66	4,50	4,64	4,56	4,56	2,33	2,26	2,36	2,37	2,64	0,23	0,21	0,14	0,16	0,16
C_18	4,70	4,62	4,46	4,61	4,52	4,52	2,29	2,22	2,32	2,33	2,60	0,35	0,14	0,41	0,19	0,19
C_19	4,70	4,62	4,46	4,60	4,51	4,52	2,29	2,22	2,31	2,33	2,59	0,32	0,10	0,37	0,16	0,16
C_20	4,65	4,57	4,41	4,56	4,47	4,48	2,24	2,18	2,27	2,29	2,55	0,29	0,08	0,34	0,18	0,13



7. З точкового графіку одразу можна спрогнозувати, що в результаті повинно бути 3 очевидних кластери.

Результати для ненормалізованих і нормалізованих даних для методу “близького сусіда” при використанні евклідової і манхеттенівської відстані вийшли дуже схожі. Чітко видно 3 кластери.

Для методу “далекого сусід”а також чітко видно три кластери, але це виглядає менш очевидно, на мою думку, ніж для методу “близького сусіда”, через те що, коли залишається два кластери, то обираються найдальші точки, нам складніше побачити, що в реальності відстані між (2,3) кластером та першим кластером не сильно відрізняється від відстані між 2 і 3 кластером, що краще показано при використанні методу “близького сусіда”.

Висновки:

В процесі виконання комп'ютерного практикуму №3 я розглянув види мір подібності, опанував методи розрахунку деяких мір відстані, набув умінь використання базових алгоритмів кластеризування та вмінь проведення кластерного аналізу сукупності об'єктів за допомогою застосунку Statistica.

Проводити кластеризування в Statistica дуже зручно, для цього, по факту, треба натиснути на 3 кнопки. І ти отримуєш широкий спектр результатів роботи від матриці відстані до того в якому порядку увійшли в який кластер елементи.