

K-MEANS ΜΕ SPSS

Αναπλ. Καθηγ. Στελιος Ζήμερας
Τμήμα Στατιστικής και Αναλογιστικών –
Χρηματοοικονομικών Μαθηματικών
Πανεπιστήμιο Αιγαίου
Σαμος

2021

Εισαγωγή

A/A	A1	A2	B1	B2	E1	E2	G1	G2	D1	D2	D3	E3
1	20.2	21.2	37.2	29.2	1382.6	332.9	19.2	17.2	29.2	19.2	22.2	1522.3
2	21.3	21.9	26.0	15.1	772.9	310.8	18.8	18.2	26.9	18.1	26.4	1608.8
3	19.9	19.9	19.3	19.3	717.3	423.2	21.0	23.6	20.0	19.7	19.2	1172.9
4	15.2	16.6	27.2	20.0	753.5	461.2	21.5	22.4	27.6	23.4	24.3	1504.3
5	11.1	12.0	23.4	16.1	583.0	397.0	14.1	13.9	23.0	12.9	19.9	1295.8
6	8.2	11.2	21.3	13.9	666.0	473.8	14.8	18.4	22.3	10.7	17.3	1252.7
7	12.5	12.0	20.5	15.1	557.5	424.3	17.4	15.9	25.0	21.0	19.2	817.8
8	9.8	9.5	21.0	12.4	415.3	436.6	16.7	16.3	25.4	18.1	17.5	1049.8
9	13.7	13.5	21.4	16.2	575.2	331.1	17.4	18.9	25.2	19.5	22.8	1078.5
10	16.6	16.4	23.8	18.7	657.9	582.6	23.9	21.0	25.3	17.2	21.1	1114.1
11	17.2	15.3	23.0	17.6	700.5	436.2	17.6	15.0	28.4	18.9	25.8	1399.7
12	26.2	14.8	16.3	18.5	735.6	496.6	17.1	18.1	28.3	15.6	27.2	1687.5
13	19.3	21.0	18.8	16.4	782.6	439.4	13.4	15.2	20.6	9.9	18.7	1100.9
14	16.3	16.8	19.4	17.3	614.8	380.1	18.5	17.4	27.0	19.1	23.2	1395.5
15	15.2	16.5	22.1	16.7	547.2	303.5	18.0	16.5	25.7	18.1	21.7	1282.6
16	12.7	13.6	20.2	15.0	583.1	329.0	17.8	16.3	27.4	13.2	24.3	1394.6
17	14.2	17.1	24.8	18.4	733.5	352.0	16.5	17.4	28.4	16.8	227.1	1513.0
18	15.6	16.5	23.3	19.2	578.0	418.8	18.4	17.5	24.6	19.3	19.9	1313.6
19	23.0	18.4	28.7	20.5	1116.7	478.5	21.8	24.0	25.7	21.3	22.9	1679.1
20	16.2	17.2	21.9	17.6	607.5	476.1	18.1	18.5	27.6	16.6	24.1	1489.1

Εισαγωγή

patientsID	A1	A2	B1	B2	E1	E2	G1	G2	D1	D2	D3	E3
1,00	202,00	212,00	372,00	292,00	13826,00	3329,00	192,00	172,00	292,00	192,00	222,00	15223,00
2,00	213,00	219,00	260,00	151,00	7729,00	3108,00	188,00	182,00	269,00	181,00	264,00	16088,00
3,00	199,00	199,00	193,00	193,00	7173,00	4232,00	210,00	236,00	200,00	197,00	192,00	11729,00
4,00	152,00	166,00	272,00	200,00	7535,00	4612,00	215,00	224,00	276,00	234,00	243,00	15043,00
5,00	111,00	120,00	234,00	161,00	5830,00	3970,00	141,00	139,00	230,00	129,00	199,00	12958,00
6,00	82,00	112,00	213,00	139,00	6660,00	4738,00	148,00	184,00	223,00	107,00	173,00	12527,00
7,00	125,00	120,00	205,00	151,00	5575,00	4243,00	174,00	159,00	250,00	210,00	192,00	8178,00
8,00	98,00	95,00	210,00	124,00	4153,00	4366,00	167,00	163,00	254,00	181,00	175,00	10498,00
9,00	137,00	135,00	214,00	162,00	5752,00	3311,00	174,00	189,00	252,00	195,00	228,00	10785,00
10,00	166,00	164,00	238,00	187,00	6579,00	5826,00	239,00	210,00	253,00	172,00	211,00	11141,00
1,00	172,00	153,00	230,00	176,00	7005,00	4362,00	176,00	150,00	284,00	189,00	258,00	13997,00
12,00	262,00	148,00	163,00	185,00	7356,00	4966,00	171,00	181,00	283,00	156,00	272,00	16875,00
13,00	193,00	210,00	188,00	164,00	7826,00	4394,00	134,00	152,00	206,00	99,00	187,00	11009,00
14,00	163,00	168,00	194,00	173,00	6148,00	3801,00	185,00	174,00	270,00	191,00	232,00	13955,00
15,00	152,00	165,00	221,00	167,00	5472,00	3035,00	180,00	165,00	257,00	181,00	217,00	12826,00
16,00	127,00	136,00	202,00	150,00	5831,00	3290,00	178,00	163,00	274,00	132,00	243,00	13946,00
17,00	142,00	171,00	248,00	184,00	7335,00	3520,00	165,00	174,00	284,00	168,00	2271,00	15130,00
18,00	156,00	165,00	233,00	192,00	5780,00	4188,00	184,00	175,00	246,00	193,00	199,00	13136,00
19,00	230,00	184,00	287,00	205,00	11167,00	4785,00	218,00	240,00	257,00	213,00	229,00	16791,00
20,00	162,00	172,00	219,00	176,00	6075,00	4761,00	181,00	185,00	276,00	166,00	241,00	14891,00

μέθοδο hierarchical clustering

Εισαγωγή

Classify → hierarchical clustering

patientsID	A1	A2	B1	B2	E1	E2	G1	G2	D1	D2	D3	E3	ve
1,00	202,00	212,00	372,00	292,00	13826,00	3329,00	192,00	172,00	292,00	192,00	222,00	15223,00	
2,00	213,00	219,00	260,00	151,00	7729,00	3108,00	188,00	182,00	269,00	181,00	264,00	16088,00	
3,00	199,00	199,00	193,00	193,00	7173,00	4232,00	210,00	236,00	200,00	197,00	192,00	11729,00	
						4612,00	215,00	224,00	276,00	234,00	243,00	15043,00	
						3970,00	141,00	139,00	230,00	129,00	199,00	12958,00	
						4728,00	148,00	184,00	223,00	107,00	173,00	12527,00	
											92,00	8178,00	
											75,00	10498,00	
											28,00	10785,00	
											11,00	11141,00	
											58,00	13997,00	
											72,00	16875,00	
											87,00	11009,00	
											32,00	13955,00	
											17,00	12826,00	
											43,00	13946,00	
											71,00	15130,00	
											99,00	13136,00	
											29,00	16791,00	
											41,00	14891,00	

Hierarchical Cluster Analysis: Method

Cluster Method: Nearest neighbor

Measure

Interval: Euclidean distance
Power: 2 Root: 2

Counts: Chi-squared measure

Binary: Squared Euclidean distance
Present: 1 Absent: 0

Transform Values

Standardize: None
 By variable
 By case

Transform Measure

Absolute values
 Change sign
 Rescale to 0-1 range

Continue Cancel Help

Hierarchical Cluster Analysis

Variables(s): G2, D1, D2, D3, E3

Label Cases by: patientsID
Cluster
 Cases Variables

Display
 Statistics Plots

Statistics... Plots... Method... Save...

OK Paste Reset Cancel Help

Μέθοδοι

Η μέθοδος (Cluster Method) που έχει επιλεγεί είναι η nearest neighbours analysis (or single linkage analysis) η οποία επιτρέπει την ταυτοποίηση ομάδων(συστάδων) αντικειμένων, στην προκειμένη περίπτωση ατόμων με κοινά μορφολογικά χαρακτηριστικά στο ιερό οστό, με βάση τις μεταβλητές που εισάγονται.

Η μέθοδος ξεκινά μετρώντας τις αποστάσεις μεταξύ των αντικειμένων, στην προκειμένη περίπτωση ατόμων. Στην αρχή, το κάθε άτομο αποτελεί μια ομάδα από μόνο του, και σταδιακά με βάση τη μετρούμενη Ευκλείδεια απόσταση (όπως φαίνεται και η επιλογή της μεθόδου μέτρησης στην αμέσως προηγούμενη εικόνα) ξεκινάει η δημιουργία μεγαλύτερων ομάδων καθώς η μετρούμενη απόσταση μεγαλώνει. Στην αρχή δηλαδή προστίθενται στην ίδια ομάδα τα δύο άτομα που απέχουν λιγότερο. Στη συνέχεια τα δύο άτομα με τη δεύτερη μεγαλύτερη απόσταση σχηματίζουν ομάδα και η διαδικασία αυτή συνεχίζεται μέχρι τελικά να τοποθετηθούν και τα 20 άτομα που περιλαμβάνονται στο συγκεκριμένο παράδειγμα. Τελικά δηλαδή, όλα τα άτομα κατατάσσονται στην ίδια ομάδα. Αυτή η διαδικασία φαίνεται απεικονιστικά στο δενδρόγραμμα.

Μέθοδοι

Στον επόμενο πίνακα (proximity ή distance matrix) φαίνονται οι υπολογιζόμενες αποστάσεις μεταξύ των ατόμων που περιλαμβάνονται στη μελέτη. Τα άτομα υπάρχουν τόσο οριζόντια όσο και κάθετα και δίνεται η αντίστοιχη τιμή της απόστασης εκεί που διασταυρώνονται. Προφανώς για το ίδιο άτομο η απόσταση είναι 0.

Case	Euclidean Distance																			
	1: 1,00	2: 2,00	3: 3,00	4: 4,00	5: 5,00	6: 6,00	7: 7,00	8: 8,00	9: 9,00	10: 10,00	11: 1,00	12: 12,00	13: 13,00	14: 14,00	15: 15,00	16: 16,00	17: 17,00	18: 18,00	19: 19,00	20: 20,00
1: 1,00	,000	6164,868	7572,429	6425,258	8339,191	7790,687	10890,917	10819,028	9216,286	8686,485	7009,730	6880,189	7413,729	7799,502	8698,624	8100,456	6812,511	8358,704	3416,463	7891,799
2: 2,00	6164,868	,000	4538,044	1845,683	3765,585	4065,565	8277,906	6757,160	5664,684	5761,630	2545,197	2056,658	5242,682	2746,022	3968,771	2871,353	2297,903	3700,384	3890,784	2628,481
3: 3,00	7572,429	4538,044	,000	3358,205	1849,099	1093,673	3897,030	3269,621	1943,057	1802,527	2283,304	5203,903	997,283	2491,009	2354,679	2762,816	4055,014	1983,234	6472,774	3391,066
4: 4,00	6425,258	1845,683	3358,205	,000	2775,059	2675,201	7150,208	5673,357	4797,236	4197,741	1204,050	1883,844	4056,518	1943,599	3416,172	2424,692	2315,857	2627,914	4035,440	1479,098
5: 5,00	8339,191	3765,585	1849,099	2775,059	,000	1212,270	4795,580	3004,762	2274,394	2707,406	1621,790	4324,848	2825,098	1067,773	1014,731	1202,509	3389,516	306,948	6624,764	2106,399
6: 6,00	7790,687	4065,565	1093,673	2675,201	1212,270	,000	4511,227	3247,814	2431,409	1771,868	1565,998	4415,445	1951,195	1790,175	2102,252	2193,129	3624,158	1212,127	6209,464	2440,217
7: 7,00	10890,917	8277,906	3897,030	7150,208	4795,580	4511,227	,000	2724,506	2774,779	3508,264	5994,198	8908,752	3624,014	5822,756	4804,022	5852,618	7502,128	4963,162	10285,071	6752,283
8: 8,00	10819,028	6757,160	3269,621	5673,357	3004,762	3247,814	2724,506	,000	1939,203	2907,479	4516,300	7164,648	3713,159	4033,158	2990,424	3984,090	6058,336	3106,846	9435,249	4813,223
9: 9,00	9216,286	5664,684	1943,057	4797,236	2274,394	2431,409	2774,779	1939,203	,000	2672,874	3605,141	6513,321	2355,836	3232,461	2079,041	3162,967	5060,382	2510,142	8221,317	4366,913
10: 10,00	8686,485	5761,630	1802,527	4197,741	2707,406	1771,868	3508,264	2907,479	2672,874	,000	3239,271	5852,139	1910,873	3494,630	3443,963	3856,526	5104,251	2703,055	7352,993	3931,508
11: 1,00	7009,730	2545,197	2283,304	1204,050	1621,790	1565,998	5994,198	4516,300	3605,141	3239,271	,000	2964,096	3103,192	1026,683	2342,153	1592,979	2481,224	1509,476	5032,696	1351,315
12: 12,00	6880,189	2056,658	5203,903	1883,844	4324,848	4415,445	8908,752	7164,648	6513,321	5852,139	2964,096	,000	5914,924	3370,079	4867,582	3706,296	3025,685	4134,480	3820,091	2373,660
13: 13,00	7413,729	5242,682	997,283	4056,518	2825,098	1951,195	3624,014	3713,159	2355,836	1910,873	3103,192	5914,924	,000	3444,808	3272,155	3721,013	4727,691	2961,839	6692,711	4276,649
14: 14,00	7799,502	2746,022	2491,009	1943,599	1067,773	1790,175	5822,756	4033,158	3232,461	3494,630	1026,683	3370,079	3444,808	,000	1523,111	606,953	2651,544	979,574	5849,961	1343,337
15: 15,00	8698,624	3968,771	2354,679	3416,172	1014,731	2102,252	4804,022	2990,424	2079,041	3443,963	2342,153	4867,582	3272,155	1523,111	,000	1205,735	3638,088	1233,638	7158,019	2758,388
16: 16,00	8100,456	2871,353	2762,816	2424,692	1202,509	2193,129	5852,618	3984,090	3162,967	3856,526	1592,979	3706,296	3721,013	606,953	1205,735	,000	2799,279	1214,962	6232,149	1766,789
17: 17,00	6812,511	2297,903	4055,014	2315,857	3389,516	3624,158	7502,128	6058,336	5060,382	5104,251	2481,224	3025,685	4727,691	2651,544	3638,088	2799,279	,000	3337,128	4820,055	2703,230
18: 18,00	8358,704	3700,384	1983,234	2627,914	306,948	1212,127	4963,162	3106,846	2510,142	2703,055	1509,476	4134,480	2961,839	979,574	1233,638	1214,962	3337,128	,000	6538,419	1870,673
19: 19,00	3416,463	3890,784	6472,774	4035,440	6624,764	6209,464	10285,071	9435,249	8221,317	7352,993	5032,696	3820,091	6692,711	5849,961	7158,019	6232,149	4820,055	6538,419	,000	5436,578
20: 20,00	7891,799	2628,481	3391,066	1479,098	2106,399	2440,217	6752,283	4813,223	4366,913	3931,508	1351,315	2373,660	4276,649	1343,337	2758,388	1766,789	2703,230	1870,673	5436,578	,000

This is a dissimilarity matrix

Μέθοδοι

Στον πίνακα που ακολουθεί, φαίνονται πιο απλοποιημένα τα στοιχεία του προηγούμενου. Δηλαδή έχει γίνει κατάταξη των ατόμων με βάση την απόσταση, ξεκινώντας από την μικρότερη. Η ελάχιστη απόσταση είναι μεταξύ των ατόμων 5 και 18. Οπότε σχηματίζεται μια ομάδα. Η αμέσως επόμενη είναι μεταξύ των ατόμων 14 και 16 και η διαδικασία αυτή συνεχίζεται μέχρι να τοποθετηθούν όλα τα άτομα, τελικά σε μια κοινή ομάδα.

Μέθοδοι

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	5	18	306,948	0	0	3
2	14	16	606,953	0	0	3
3	5	14	979,574	1	2	5
4	3	13	997,283	0	0	7
5	5	15	1014,731	3	0	6
6	5	11	1026,683	5	0	8
7	3	6	1093,673	4	0	9
8	4	5	1204,050	0	6	9
9	3	4	1212,127	7	8	10
11	3	10	1771,868	10	0	12
12	2	3	1845,683	0	11	13
13	2	12	1883,844	12	0	15
14	8	9	1939,203	0	0	15
15	2	8	1943,057	13	14	16
16	2	17	2297,903	15	0	17
17	2	7	2724,506	16	0	19
18	1	19	3416,463	0	0	19
19	1	2	3820,091	18	17	0

Μέθοδοι

Στη συνέχεια φαίνεται εποπτικά η προηγούμενη διαδικασία. Στον κάθετο άξονα είναι τοποθετημένα τα άτομα και στον οριζόντιο οι αποστάσεις. Όπως παρατηρείται στην αρχή σχηματίζει ομάδα το άτομο 5 με το άτομο 18, στη συνέχεια το 14 με το 16. Έπειτα και τα τέσσερα αυτά άτομα κάνουν όλα μαζί μια ομάδα. Στη συνέχεια το 3 με το 13 και σε επόμενο στάδιο το 15 με το 11 σχηματίζουν ομάδα. Καθώς αυξάνεται η απόσταση η ομάδα (15,11) ενώνεται με αυτή των (5, 18, 14, 16) κ.ο.κ. Τελικά φαίνονται δύο ομάδες μία μεταξύ των 1 και 19 και μία μεταξύ των υπολοίπων οι οποίες στο τέλος ενώνονται σε μια και μοναδική. Το δεντροδιαγραμμα παρακάτω βοηθάει στην κατανόηση υλοποίησης του αλγόριθμου.

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	5	18	306,948	0	0	3
2	14	16	606,953	0	0	3
3	5	14	979,574	1	2	5
4	3	13	997,283	0	0	7
5	5	15	1014,731	3	0	6
6	5	11	1026,683	5	0	8
7	3	6	1093,673	4	0	9
8	4	5	1204,050	0	6	9
9	3	4	1212,127	7	8	10

Μέθοδοι

Dendrogram using Single Linkage
Rescaled Distance Cluster Combine

