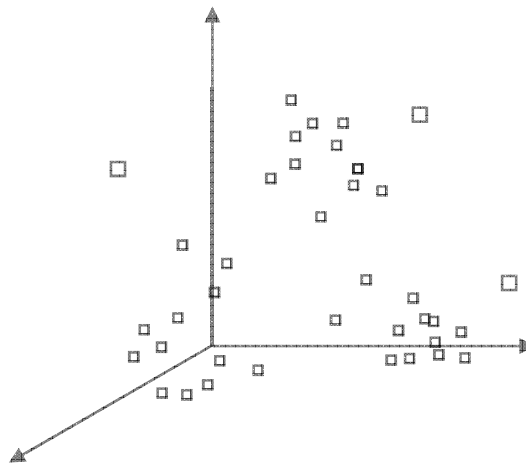


Εφαρμογές της Παραγοντικής Ανάλυσης των Αντιστοιχιών



Άγγελος Μάρκος
Τμήμα Εφαρμοσμένης Πληροφορικής
Πανεπιστήμιο Μακεδονίας

Περιεχόμενο Παρουσίασης

- ❑ Η θέση της Παραγοντικής Ανάλυσης των Αντιστοιχιών (Correspondence Analysis - CA) στο Χάρτη των Πολυμεταβλητών Μεθόδων

- ❑ 2 + 2 Εφαρμογές της CA
 - ❑ Αντιλήψεις για την Υγεία (1)

 - ❑ Αντιλήψεις για την Υγεία (2)

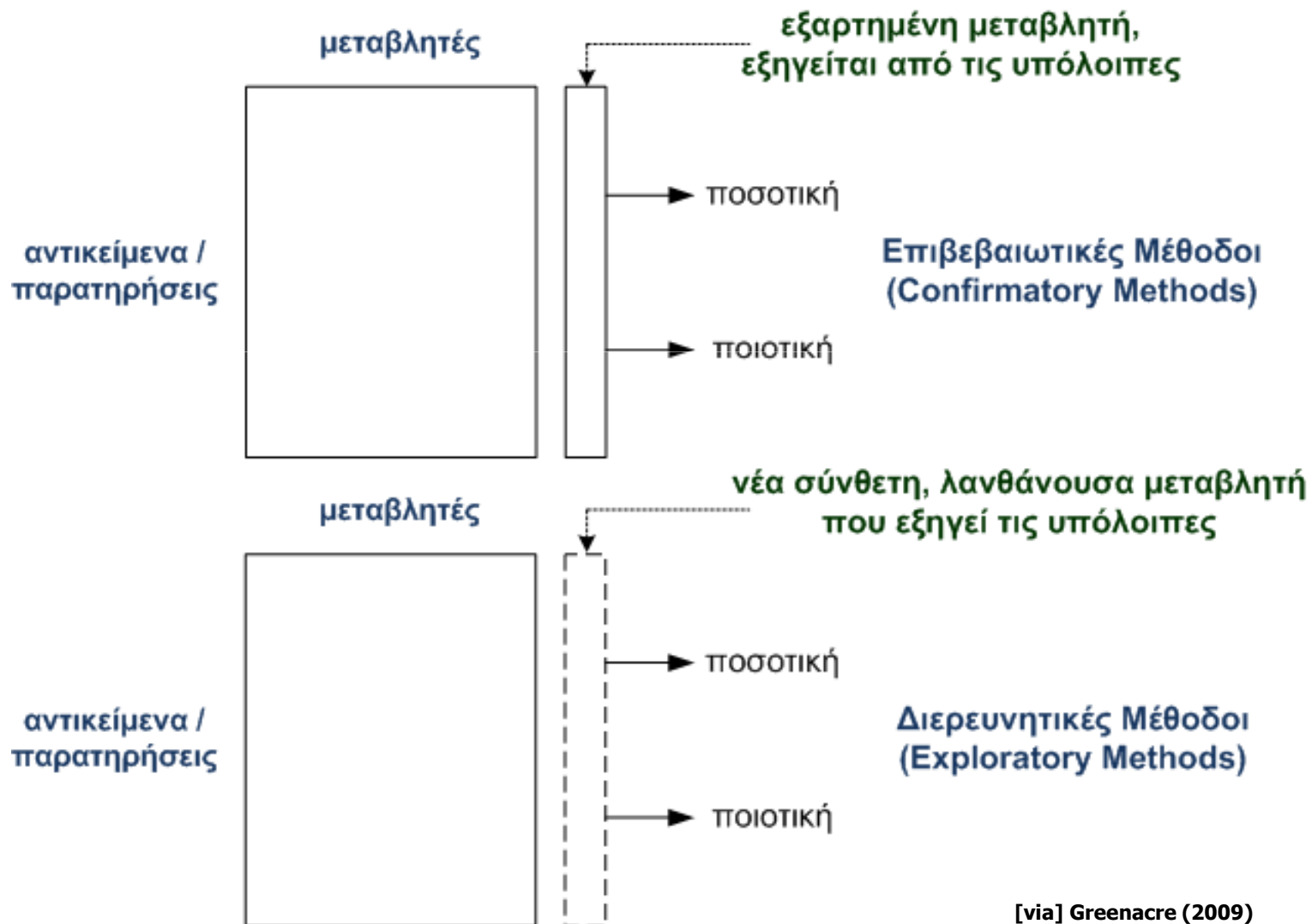
 - ❑ Αντιλήψεις για την Επιστήμη (1)

 - ❑ Αντιλήψεις για την Επιστήμη (2)

- ❑ Το Λογισμικό της CA

Πολυμεταβλητή Ανάλυση Δεδομένων

Οι μέθοδοι της Πολυμεταβλητής Ανάλυσης Δεδομένων διακρίνονται σε δύο μεγάλες κατηγορίες, ως προς το είδος των μεταβλητών και το ζητούμενο της ερευνητικής εργασίας.



[via] Greenacre (2009)

Διερευνητικές Μέθοδοι

Μέθοδοι που αποκαλύπτουν συνεχείς δομές
(παράγοντες, κλίμακες, διαστάσεις...)

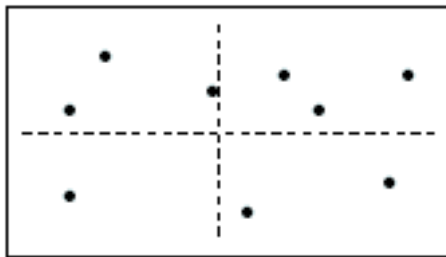
Μέθοδοι που αποκαλύπτουν διακριτές δομές
(συστάδες, ομάδες, διαμερίσεις...)

Παραγοντικές μέθοδοι

Παραγ. Ανάλυση Αντιστοιχιών (CA)
Ανάλυση σε Κύριες Συνιστώσες (PCA)
Ανάλυση Παραγόντων (FA)

Μέθοδοι Κλιμακοποίησης

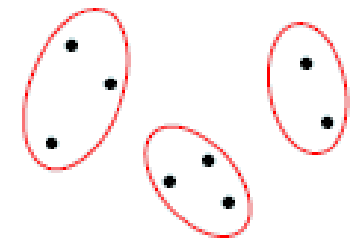
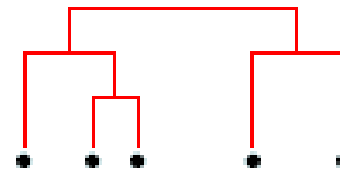
Κλασική Κλιμακοποίηση (CS)
Πολυδιάστατη Κλιμακοποίηση (MDS)



Ανάλυση Συστάδων

Ιεραρχικές
HCA

Μη Ιεραρχικές
k-means



Βασική Έννοια: **Απόσταση**

Η Παραγοντική Ανάλυση των Αντιστοιχιών

□ Χρησιμοποιείται για τη διερεύνηση της σχέσης μεταξύ δύο ή περισσότερων κατηγορικών μεταβλητών, χωρίς τη διάκριση αυτών σε εξαρτημένες και ανεξάρτητες και χωρίς αυστηρές a priori παραδοχές σχετικά με τη θεωρητική κατανομή που ακολουθούν τα δεδομένα και οι παράμετροι του υπό μελέτη πληθυσμού ή πληθυσμών.

□ Κατάλληλοι Πίνακες Εισόδου

Διμεταβλητή περίπτωση:

- Πίνακας συνάφειας δύο ή περισσότερων μεταβλητών με μη αρνητικά στοιχεία.

Πολυμεταβλητή περίπτωση:

- Λογικός πίνακας (0-1) ή γενικευμένος πίνακας συμπτώσεων (Burt).

□ Βρίσκει εφαρμογές σε όλο σχεδόν το φάσμα των επιστημονικών πεδίων.

Εφαρμογή 1. Απλός Πίνακας Συνάφειας Δύο Μεταβλητών

Δεδομένα: Αντιλήψεις για την Υγεία

Πίνακας 1. Προσωπική Αντίληψη της Κατάστασης Υγείας 6371 Ατόμων

Κατάσταση Υγείας	<i>Πολύ Καλή</i>			<i>Πολύ Κακή</i>	
Ηλικία	<i>Πολύ Καλή</i>	<i>Καλή</i>	<i>Μέτρια</i>	<i>Κακή</i>	<i>Κακή</i>
16-24	243	789	167	18	6
25-34	220	809	164	35	6
35-44	147	658	181	41	8
45-54	90	469	236	50	16
55-64	53	414	306	106	30
65-74	44	267	284	98	20
75+	20	136	157	66	17

Πηγή: Spanish National Health Survey, 1997

Ερευνητικό Πρόβλημα: Η διερεύνηση της σχέσης ανάμεσα στην ηλικία και την προσωπική αντίληψη της κατάστασης υγείας.

Εφαρμογή 1. Απλός Πίνακας Συνάφειας Δύο Μεταβλητών

Δεδομένα: Αντιλήψεις για την Υγεία

Πίνακας 1. Προσωπική Αντίληψη της Κατάστασης Υγείας 6371 Ατόμων

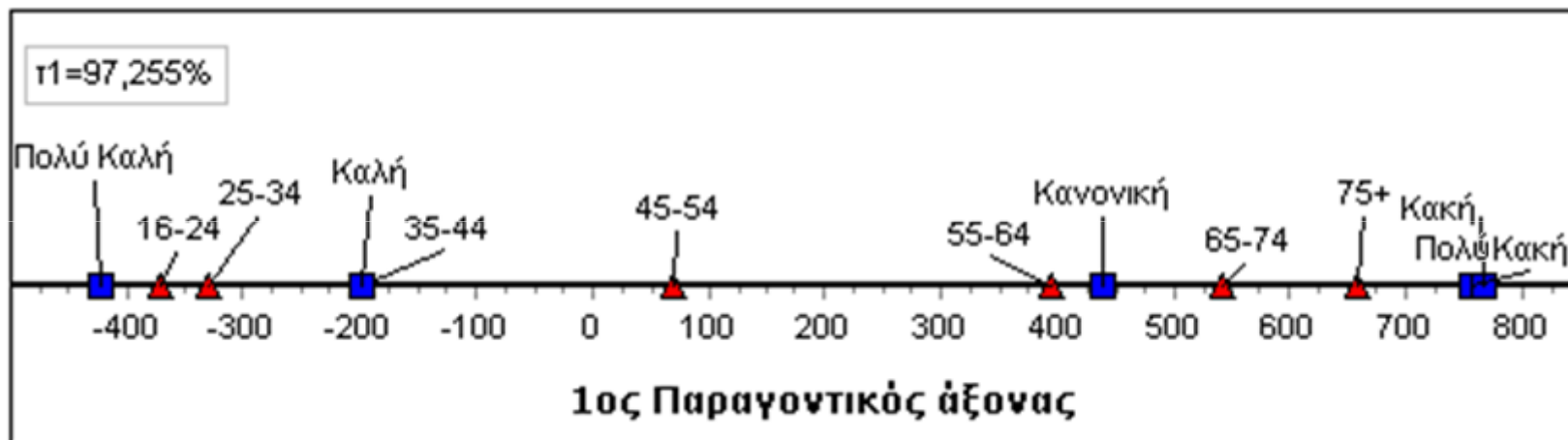
Κατάσταση Υγείας	<i>Πολύ Καλή</i>	<i>Καλή</i>	<i>Μέτρια</i>	<i>Κακή</i>	<i>Πολύ Κακή</i>
Ηλικία					
16-24	243	789	167	18	6
25-34	220	809	164	35	6
35-44	147	658	181	41	8
45-54	90	469	236	50	16
55-64	53	414	306	106	30
65-74	44	267	284	98	20
75+	20	136	157	66	17

Πηγή: Spanish National Health Survey, 1997

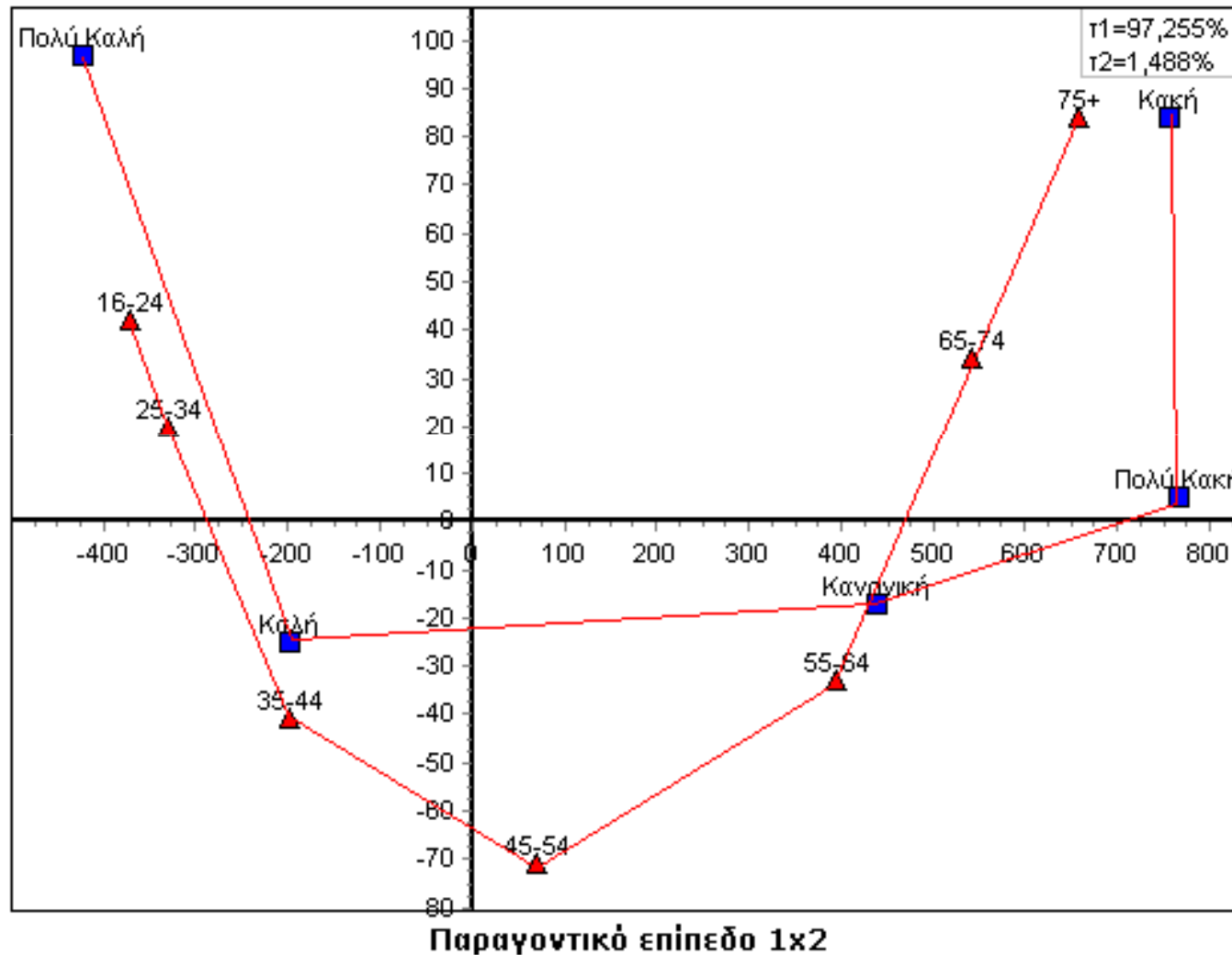
$$\chi^2 = 894,861, df = 24, p < 0,000$$

Ο στατιστικός έλεγχος χ^2 δείχνει της ύπαρξη σχέσης ανάμεσα στις δύο μεταβλητές. Που όμως οφείλεται αυτή η σχέση;

Εφαρμογή 1. Απλός Πίνακας Συνάφειας Δύο Μεταβλητών



Εφαρμογή 1. Απλός Πίνακας Συνάφειας Δύο Μεταβλητών



Εφαρμογή 2. Πίνακας Συνάφειας με Τρεις Μεταβλητές

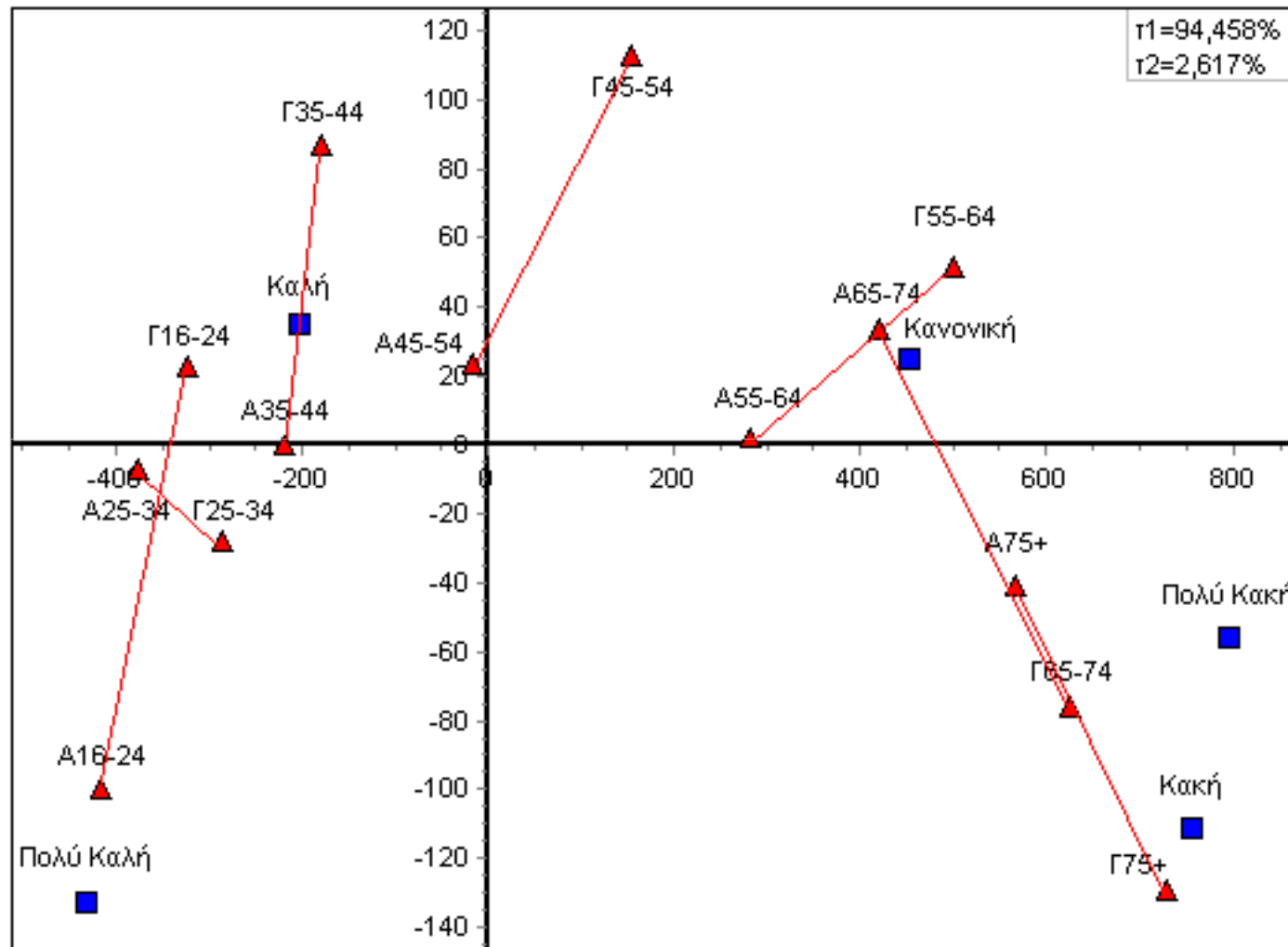
Πίνακας 2. Προσωπική Αντίληψη της Κατάστασης Υγείας 6371 Αντρών και Γυναικών

Κατάσταση Υγείας	<i>Πολύ Καλή</i>		<i>Μέτρια</i>		<i>Πολύ Κακή</i>	
Ηλικία	<i>Καλή</i>	<i>Καλή</i>	<i>Μέτρια</i>	<i>Κακή</i>	<i>Κακή</i>	<i>Καλή</i>
Άντρες						
16-24	145	402	84	5	3	639
25-34	112	414	74	13	2	615
35-44	80	331	82	24	4	521
45-54	54	231	102	22	6	415
55-64	30	219	119	53	12	433
65-74	18	125	110	35	4	292
75+	9	67	65	25	8	174
Γυναίκες						
16-24	98	387	83	13	3	584
25-34	108	395	90	22	4	619
35-44	67	327	99	17	4	514
45-54	36	238	134	28	10	446
55-64	23	195	187	53	18	476
65-74	26	142	174	63	16	421
75+	11	69	92	41	9	222

Πηγή: Spanish National Health Survey, 1997

Ερευνητικό Πρόβλημα: Η διερεύνηση της σχέσης ανάμεσα στην ηλικία, το φύλο και την προσωπική αντίληψη της κατάστασης υγείας.

Εφαρμογή 2. Πίνακας Συνάφειας με Τρεις Μεταβλητές



Παραγοντικό επίπεδο 1x2

Εφαρμογή 3. Πίνακας «αντικείμενα x μεταβλητές»

Δεδομένα: Αντιλήψεις για την Επιστήμη

Πίνακας 3. Αντιλήψεις για την Επιστήμη

	A	B	C	D	Φύλο	Ηλικία	Μόρφωση
1	2	3	4	3	2	2	3
2	3	4	2	3	1	3	4
3	2	3	2	4	2	3	2
4	2	2	2	2	1	2	3
5	3	3	3	3	1	5	2
...
870	3	4	2	3	1	2	2
871	1	2	2	2	2	3	6

Πηγή: International Social Science Survey, 1994 (Γερμανία)

Κλίμακα

1. Συμφωνώ απόλυτα
2. Συμφωνώ
3. Ούτε συμφωνώ / ούτε διαφωνώ
4. Διαφωνώ
5. Διαφωνώ απόλυτα

Σε ποιο βαθμό συμφωνείτε ή διαφωνείτε με τις παρακάτω προτάσεις;

- A.** Εμπιστευόμαστε υπερβολικά την επιστήμη κι όχι αρκετά (όσο θα έπρεπε) την θρησκευτική πίστη.
- B.** Γενικά, η επιστήμη σήμερα κάνει περισσότερο κακό παρά καλό.
- C.** Η οποιαδήποτε αλλαγή προκαλεί ο άνθρωπος στη φύση – ασχέτως με το πόσο επιστημονική είναι ή όχι – θα χειροτερέψει τα πράγματα.
- D.** Η σύγχρονη επιστήμη θα λύσει τα περιβαλλοντικά μας προβλήματα με μικρές αλλαγές στον τρόπο ζωής μας.

Εφαρμογή 3. Πίνακας «αντικείμενα x μεταβλητές»

Δεδομένα: Αντιλήψεις για την Επιστήμη

Πίνακας 3. Αντιλήψεις για την Επιστήμη

	A	B	C	D	Φύλο	Ηλικία	Μόρφωση
1	2	3	4	3	2	2	3
2	3	4	2	3	1	3	4
3	2	3	2	4	2	3	2
4	2	2	2	2	1	2	3
5	3	3	3	3	1	5	2
...
870	3	4	2	3	1	2	2
871	1	2	2	2	2	3	6

Πηγή: International Social Science Survey, 1994 (Γερμανία)

Κλίμακα

1. Συμφωνώ απόλυτα
2. Συμφωνώ
3. Ούτε συμφωνώ / ούτε διαφωνώ
4. Διαφωνώ
5. Διαφωνώ απόλυτα

Δημογραφικά Χαρακτηριστικά

Φύλο: άντρας, γυναίκα

Ηλικία: 16–24, 25–34, 35–44, 45–54, 55–64, 65 και άνω

Μόρφωση: Μερικές τάξεις Δημοτικού, Δημοτικό, Μερικές τάξεις Γυμνασίου ή Λυκείου, Λύκειο, Κολλέγιο, Πανεπιστήμιο

Εφαρμογή 3. Πίνακας Burt

Δεδομένα: Αντιλήψεις για την Επιστήμη

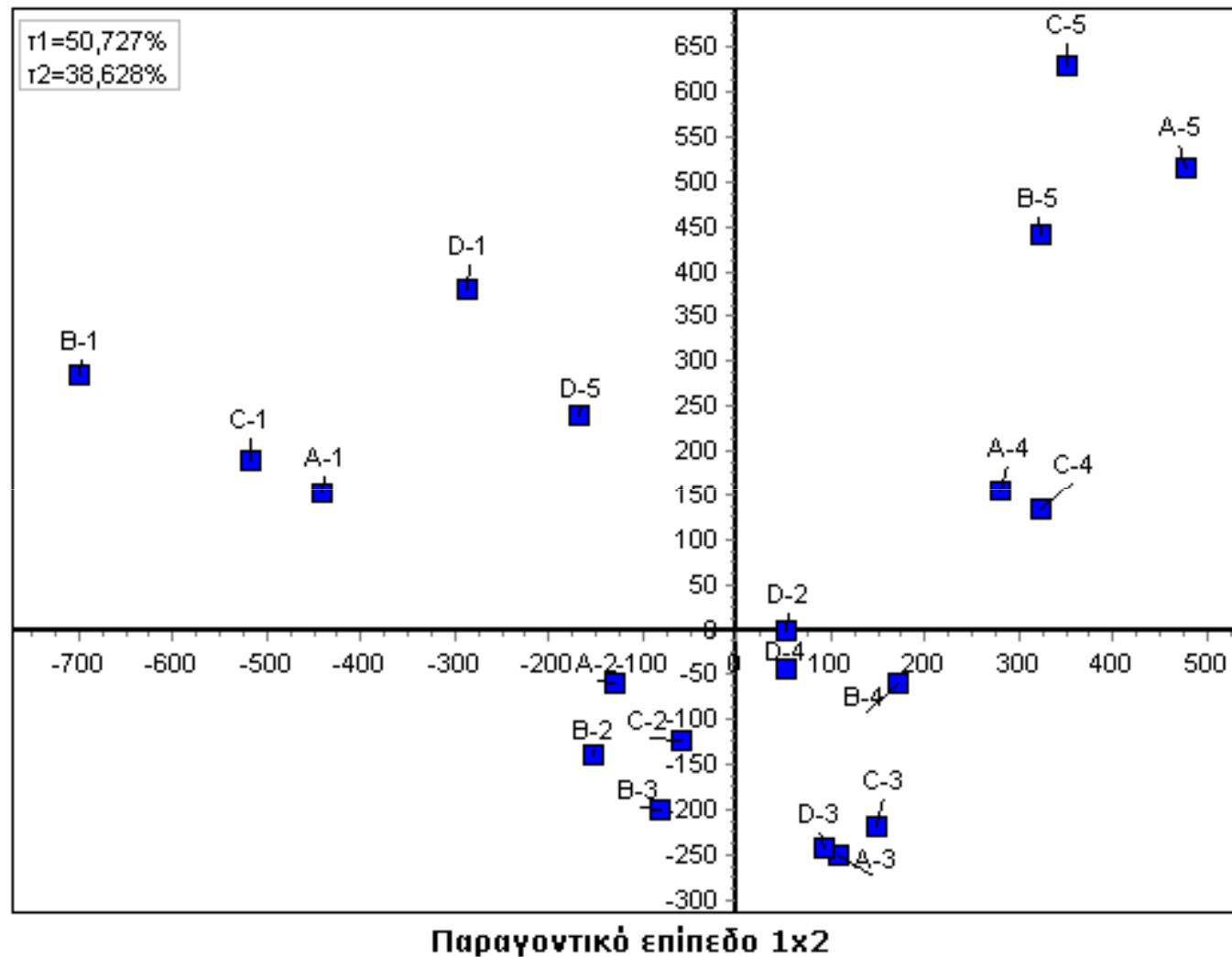
Πίνακας 4. Ο Πίνακας Burt για τα Χαρακτηριστικά A έως D

	Variable A					Variable B					Variable C					Variable D				
	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5
A1	119	0	0	0	0	27	28	30	22	12	49	40	18	7	5	15	25	17	34	28
A2	0	322	0	0	0	38	74	84	96	30	67	142	60	41	12	22	102	76	68	54
A3	0	0	204	0	0	3	48	63	73	17	18	75	70	34	7	10	44	68	58	24
A4	0	0	0	178	0	3	21	23	79	52	16	50	40	56	16	9	52	28	54	35
A5	0	0	0	0	48	0	3	5	11	29	2	9	9	16	12	4	9	13	12	10
B1	27	38	3	3	0	71	0	0	0	0	43	19	4	3	2	9	17	10	10	25
B2	28	74	48	21	3	0	174	0	0	0	36	88	34	15	1	16	51	42	45	20
B3	30	84	63	23	5	0	0	205	0	0	37	90	57	19	2	10	53	63	51	28
B4	22	96	73	79	11	0	0	0	281	0	27	88	75	74	17	6	66	70	92	47
B5	12	30	17	52	29	0	0	0	0	140	9	31	27	43	30	19	45	17	28	31
C1	49	67	18	16	2	43	36	37	27	9	152	0	0	0	0	25	24	15	38	50
C2	40	142	75	50	9	19	88	90	88	31	0	316	0	0	0	15	97	67	89	48
C3	18	60	70	40	9	4	34	57	75	27	0	0	197	0	0	5	51	83	41	17
C4	7	41	34	56	16	3	15	19	74	43	0	0	0	154	0	6	44	30	51	23
C5	5	12	7	16	12	2	1	2	17	30	0	0	0	0	52	9	16	7	7	13
D1	15	22	10	9	4	9	16	10	6	19	25	15	5	6	9	60	0	0	0	0
D2	25	102	44	52	9	17	51	53	66	45	24	97	51	44	16	0	232	0	0	0
D3	17	76	68	28	13	10	42	63	70	17	15	67	83	30	7	0	0	202	0	0
D4	34	68	58	54	12	10	45	51	92	28	38	89	41	51	7	0	0	0	226	0
D5	28	54	24	35	10	25	20	28	47	31	50	48	17	23	13	0	0	0	0	151

Πηγή: Greenacre & Blasius (2006)

Ερευνητικό Πρόβλημα: Η διερεύνηση της σχέσης ανάμεσα στα χαρακτηριστικά A έως D (στάσεις και αντιλήψεις για την επιστήμη).

Εφαρμογή 3. Πίνακας Burt



1^{ος} άξονας - Cronbach's α: 0,605

2^{ος} άξονας - Cronbach's α: 0,565

Εφαρμογή 4. Πίνακας Συνάφειας Επτά Μεταβλητών

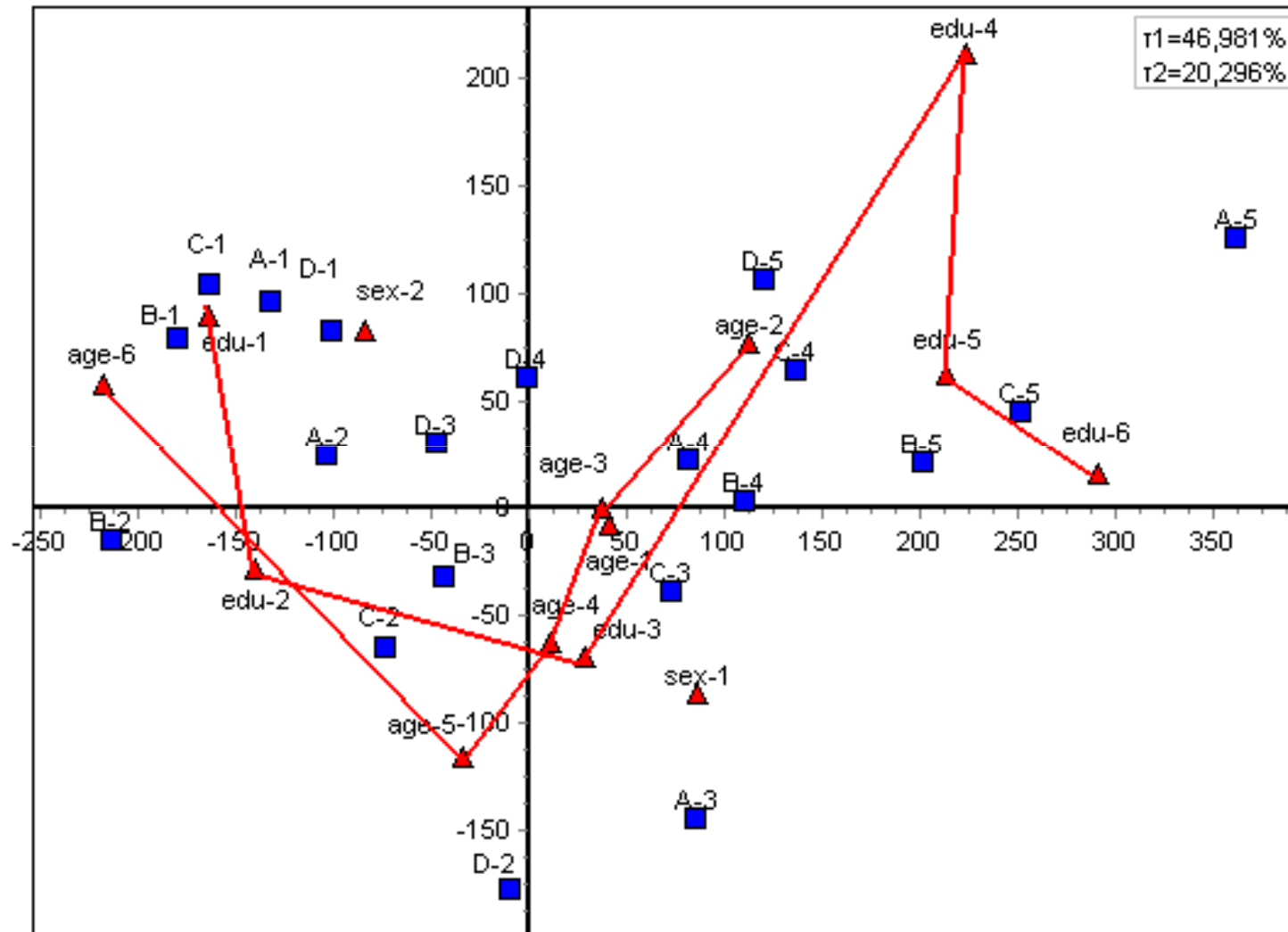
Πίνακας 5. Ο Πίνακας Συνάφειας A-D ως προς Δημογραφικά Χαρακτηριστικά

	Variable A					Variable B					Variable C					Variable D				
	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5
Sex																				
Male	43	144	120	92	28	25	75	104	146	77	58	157	105	82	25	25	136	92	101	73
Female	76	178	84	86	20	46	99	101	135	63	94	159	92	72	27	35	96	110	125	78
Age *																				
age1	9	34	17	25	6	5	21	16	33	16	13	32	22	19	5	7	26	19	30	9
age2	33	68	45	51	13	15	30	52	76	37	35	60	57	43	15	10	46	48	56	50
age3	18	68	40	26	11	15	27	31	62	23	24	61	35	25	13	12	39	42	35	30
age4	19	45	47	27	8	13	29	34	41	29	24	57	33	26	6	10	39	38	38	21
age5	18	43	33	25	5	10	26	35	37	16	21	45	26	24	8	9	48	23	28	16
age6	22	69	22	24	5	13	41	37	32	19	35	61	24	17	5	12	34	32	39	25
Education ¹																				
edu1	7	15	7	8	1	6	10	11	7	4	7	8	11	12	0	5	9	12	8	4
edu2	59	155	84	68	12	34	93	95	112	44	79	156	73	54	16	28	110	90	95	55
edu3	29	84	65	54	10	19	47	55	82	39	34	90	68	36	14	11	69	58	63	41
edu4	11	27	18	26	12	6	12	18	37	21	18	24	18	28	6	7	13	23	29	22
edu5	5	20	11	8	5	4	5	11	16	13	6	14	14	9	6	4	12	7	14	12
edu6	8	21	19	14	8	2	7	15	27	19	8	24	13	15	10	5	19	12	17	17

Πηγή: Greenacre & Blasius (2006)

Ερευνητικό Πρόβλημα: Η διερεύνηση της σχέσης ανάμεσα στα χαρακτηριστικά A έως D (στάσεις και αντιλήψεις για την επιστήμη) και τα δημογραφικά στοιχεία (Φύλο, Ηλικία, Επίπεδο Μόρφωσης).

Εφαρμογή 4. Πίνακας Συνάφειας Επτά Μεταβλητών



Παραγοντικό επίπεδο 1x2

Το Λογισμικό της CA (1)

✓ Στο SPSS:

Διμεταβλητή περίπτωση

(Analyze → Data Reduction... → Correspondence Analysis)

Πολυμεταβλητή περίπτωση

(Analyze → Data Reduction... → Optimal Scaling (Multiple CA))

✓ Στο Excel:

Στα πρόσθετα AFC97, XLSTAT

✓ Στη συναρτησιακή γλώσσα προγραμματισμού R:

Πακέτα MASS, ca, homals, anacor, factoMineR

✓ Στα πακέτα SAS, Minitab, SPAD, Statistica, JMP, ViSta κ.ά.

Το Λογισμικό της CA (2)

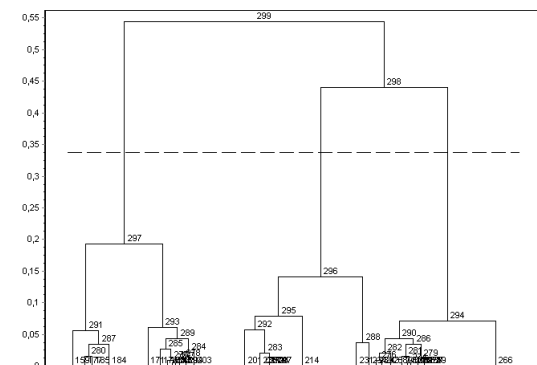
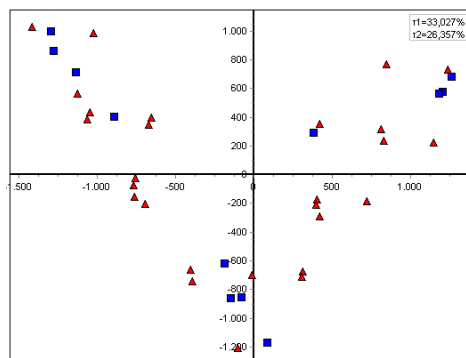
✓ Στο λογισμικό CHIC Analysis (Markos et al. 2009):

ο Συνδυάζει τους σημαντικότερους δείκτες και δυνατότητες άλλων στατιστικών πακέτων και συνδέει τις δύο βασικές μεθοδολογικές προσεγγίσεις της CA.

ο Δίνει έμφαση στην ερμηνεία των αποτελεσμάτων και στην κατασκευή ειδικών πινάκων εισόδου.

ο Διατίθεται δωρεάν.

<http://www.amarkos.gr/research/chic>



Ενδεικτική Βιβλιογραφία (1)

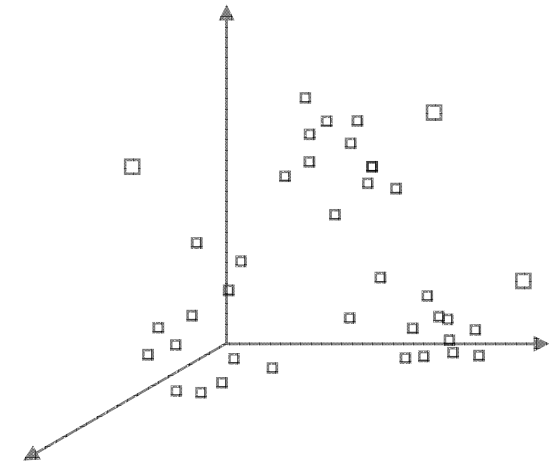
Ξενόγλωσση

- Blasius, J. & Greenacre, M.J. (2006). Multiple Correspondence Analysis and Related Methods, London: Chapman and Hall.
- Clausen, S.-E. (1998). Applied Correspondence Analysis: An Introduction. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-121, Thousand Oakes, CA: Sage.
- Gifi, A. (1996). Non-Linear Multivariate Analysis. Chichester: John Willey & Sons Ltd.
- Greenacre, M.J. (1984). Theory and Applications of Correspondence Analysis. London: Academic Press.
- Greenacre, M.J. (1993, 2007). Correspondence Analysis in Practice. London: Academic Press.
- Le Roux. B. & Rouanet, H. (2004). Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis. Dordrecht: Kluwer Academic Publishers.

Ενδεικτική Βιβλιογραφία (2)

Ελληνόγλωσση

- Αθανασιάδης, Η. (1995). Παραγοντική Ανάλυση Αντιστοιχιών και Ιεραρχική Ταξινόμηση. Εκδόσεις Νέων Τεχνολογιών, Αθήνα.
- Δρόσος, Γ. (2006). Στατιστική & Ανάλυση Δεδομένων. Εκδόσεις Ανικούλα, Θεσσαλονίκη.
- Καραπιστόλης, Δ. (1999). Ανάλυση Δεδομένων και Έρευνα Αγοράς. Εκδόσεις Ανικούλα, Θεσσαλονίκη.
- Μαυρομάτης, Γ. (1999). Στατιστικά Μοντέλα και Μέθοδοι Ανάλυσης Δεδομένων. University Studio Press, Θεσσαλονίκη.
- Μπεχράκης, Θ. (1999). Πολυδιάστατη Ανάλυση Δεδομένων: Μέθοδοι και Εφαρμογές. Εκδόσεις Νέα Σύνορα – Α.Α. Λιβάνης, Αθήνα.
- Παπαδημητρίου, Γ. (2007). Η Ανάλυση Δεδομένων. Εκδόσεις τυπωθήτω. Αθήνα.



Ευχαριστώ για την προσοχή σας

