



Η Ανάλυση Παραγόντων (Factor Analysis)

Τι είναι η ανάλυση παραγόντων



- Σκοπός της ανάλυσης παραγόντων (ΑΠ) είναι να **συνοψίσει** τις σχέσεις ανάμεσα σε ένα μεγάλο αριθμό μεταβλητών με έναν περιεκτικό και ακριβή τρόπο, ώστε να βοηθήσει **να γίνει αντιληπτή μια έννοια ή ιδιότητα**

Παράδειγμα:

- Η ΑΠ μπορεί να βοηθήσει έναν ερευνητή να αντιληφθεί ότι μια **συστοιχία από είκοσι τέσσερα επιμέρους τεστ** αντιπροσωπεύει μόνο **τέσσερις βασικές μεταβλητές**, που ονομάζονται **παράγοντες**



Είδη ανάλυσης παραγόντων

- **Διερευνητική Ανάλυση Παραγόντων**
(Exploratory Factor Analysis)
- **Επιβεβαιωτική Ανάλυση Παραγόντων**
(Confirmatory Factor Analysis)

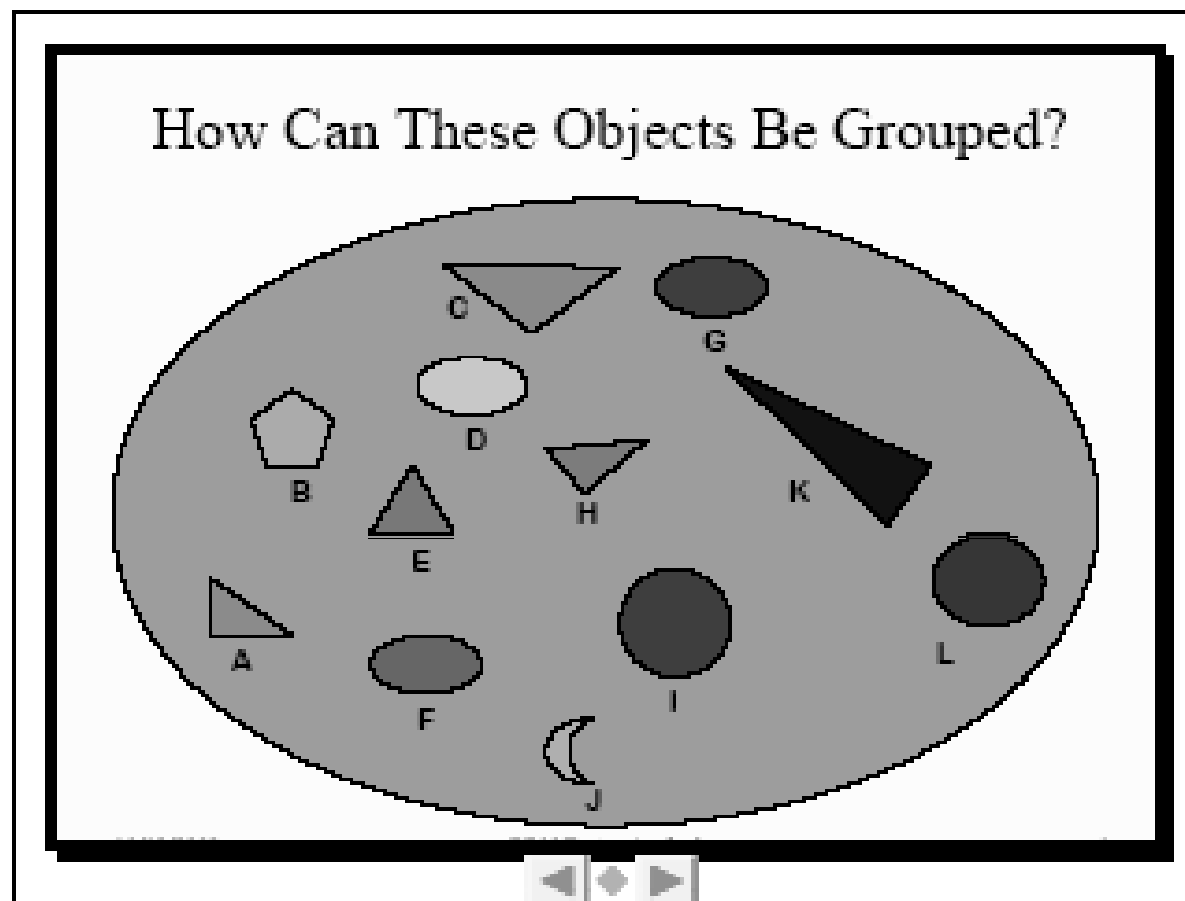
Διερευνητική Ανάλυση Παραγόντων



- Χρησιμοποιείται για την αρχική **διερεύνηση** και τη συνοπτική περιγραφή ενός σετ μεταβλητών μέσα από την **ομαδοποίησή τους**
- Δημιουργήθηκε από τον **Charles Spearman** στις αρχές του προηγούμενου αιώνα

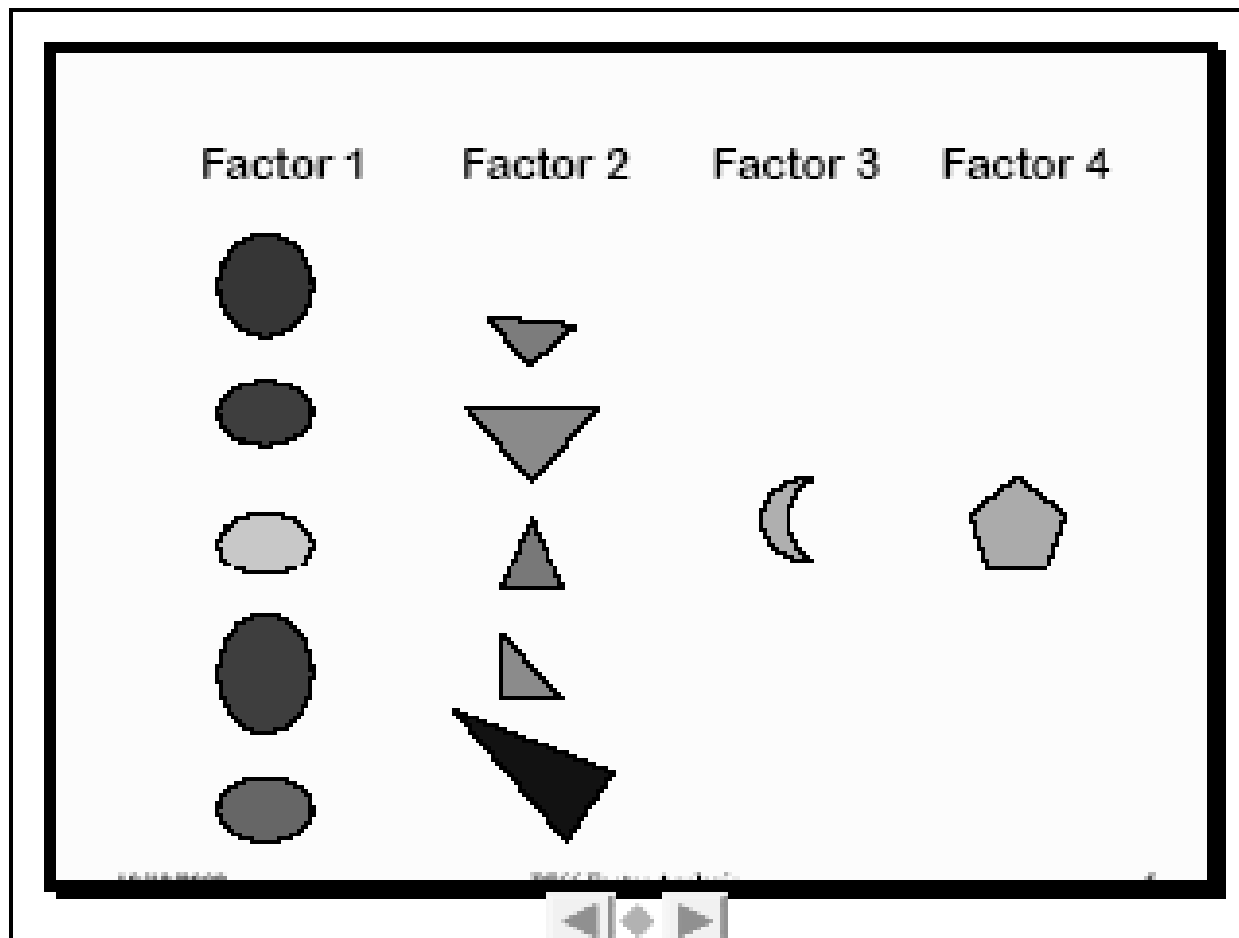


Παράδειγμα





Παράδειγμα



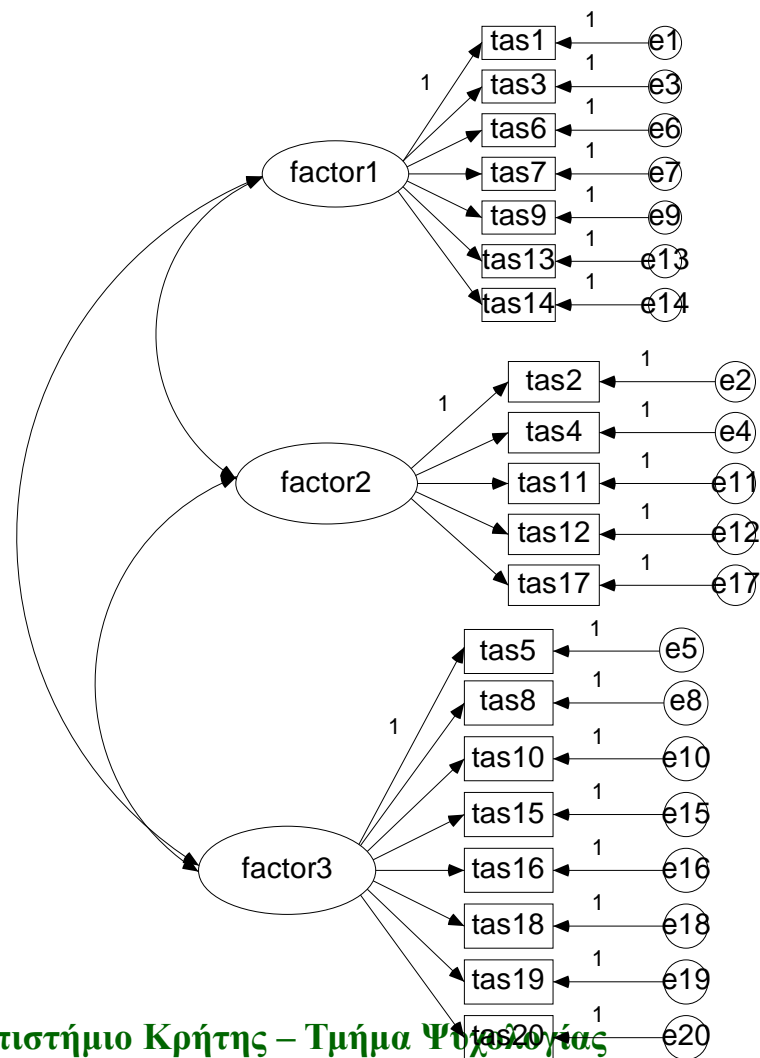
Επιβεβαιωτική Ανάλυση Παραγόντων



- Χρησιμοποιείται για να διαπιστώσει κατά πόσο ένα **προκαθορισμένο πλαίσιο σχέσεων** ανάμεσα σε κάποιες μεταβλητές (σχέσεις) **επιβεβαιώνεται** και στην πράξη (από τα δεδομένα)



Παράδειγμα



Βήματα στην Ανάλυση Παραγόντων



- Επιλέγουμε και μετράμε ένα **σετ μεταβλητών**
- Δημιουργούμε ένα **πίνακα ενδοσυναφειών** (correlation matrix)
- Επιλέγουμε τη μέθοδο **«εξαγωγής»** των παραγόντων
- Επιλέγουμε τη μέθοδο **περιστροφής** των παραγόντων
- **Ερμηνεύουμε** τους παράγοντες που προκύπτουν



Η Ποιότητα των δεδομένων

- Η Ανάλυση Παραγόντων επηρεάζεται σε σημαντικό βαθμό από **την ποιότητα** των δεδομένων που έχουμε στη διάθεσή μας
 - Οι μεταβλητές θα πρέπει να **συσχετίζονται επαρκώς** ($r > .20$)
 - Αλλά δεν πρέπει να συσχετίζονται υπερβολικά ($r < .80$)
 - Θα πρέπει οι σχέσεις να είναι ευθύγραμμες, δεν θα πρέπει να υπάρχουν ακραίες τιμές



Η Ποιότητα των δεδομένων

- Οι μεταβλητές θα πρέπει να έχουν μετρηθεί τουλάχιστον σε **κλίμακα ίσων διαστημάτων**
- Ο συνολικός αριθμός των μεταβλητών που θα αναλύσουμε θα πρέπει να είναι **3 με 5 φορές περισσότερες** από τους υποτιθέμενους παράγοντες
- Ο συνολικός αριθμός των ατόμων θα πρέπει να είναι σημαντικός (τουλάχιστον > 300)
- Θα πρέπει να υπάρχει μια αναλογία ανάμεσα στον αριθμό των μεταβλητών και των ατόμων που θα χρησιμοποιήσουμε (**10:1**, ή **5:1**)



Ποιότητα των δεδομένων

- Το SPSS παρέχει δύο δείκτες για τον έλεγχο της ποιότητας των δεδομένων
 - Ο Δείκτης **Keiser-Meyer-Olkin** αξιολογεί την επάρκεια του δείγματος (**>.50**)
 - Ο Δείκτης **Bartlett's Test of Sphericity** αξιολογεί το κατά πόσο οι συσχετίσεις μεταξύ των μεταβλητών επιτρέπουν την εφαρμογή της ανάλυσης παραγόντων (**p <0.05**)



Ποιότητα των δεδομένων

- Οι Δείκτες **Keiser-Meyer-Olkin** και **Bartlett's Test of Sphericity** όπως εμφανίζονται στο output του **SPSS**

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,951
Bartlett's Test of Sphericity	Approx. Chi-Square	44323,516
	df	1540
	Sig.	,000



Ο πίνακας ενδοσυναφειών (Correlation Matrix)

1.00				
0.87	1.00			
0.04	0.11	1.00		
0.06	0.10	0.51	1.00	
0.14	0.08	0.61	0.49	1.00

Η «εξαγωγή» των παραγόντων



- Είναι η μαθηματική μέθοδος με την οποία οι παράγοντες **καθορίζονται** από ένα μεγαλύτερο αριθμό μεταβλητών (π.χ. στοιχείων ενός τεστ)

Η μέθοδος «εξαγωγής» των παραγόντων



- **Ανάλυση Παραγόντων (Factor Analysis)**

ή

- **Ανάλυση Κυρίων Συνιστωσών; (Principal Components Analysis)**

Ανάλυση Κυρίων Συνιστωσών (Principal Components Analysis)



- Ο στόχος είναι να **μελετηθεί όλη η υπάρχουσα διακύμανση** (κοινή, μοναδική και σφάλμα) ώστε να «εξαχθεί» το μεγαλύτερο ποσοστό της διακύμανσης από τους **λιγότερους δυνατούς παράγοντες**.
 - Παράγει **συνιστώσες (components)**
 - Είναι καλύτερη μέθοδος όταν θέλουμε να **μειώσουμε τον αριθμό** των μεταβλητών

Ανάλυση Παραγόντων (Factor Analysis)

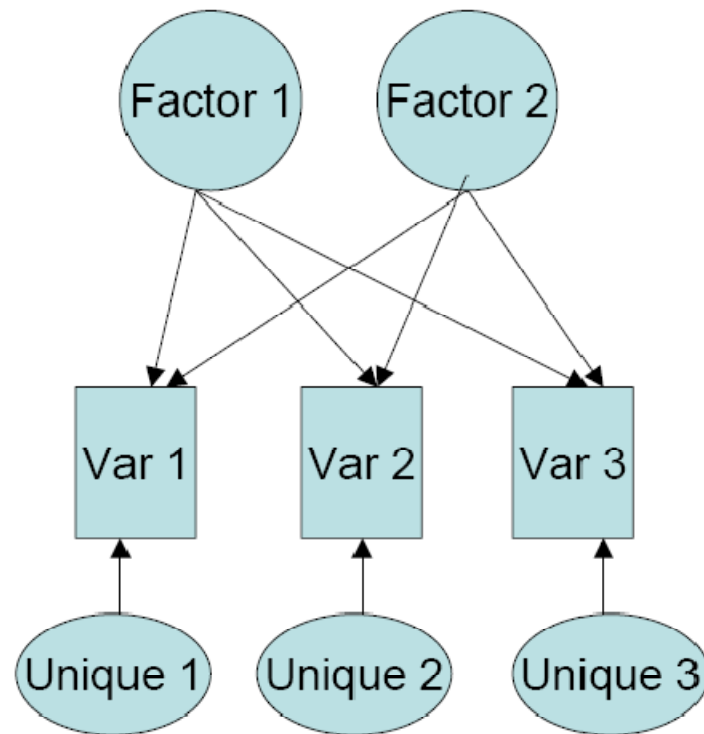


- Ο στόχος είναι να μελετηθεί **μόνο** το ποσοστό της διακύμανσης το οποίο έχουν **κοινό** οι μεταβλητές που μελετάμε
 - Παράγει **παράγοντες (factors)**
 - Είναι καλύτερη όταν θέλουμε να **κατασκευάσουμε παράγοντες**

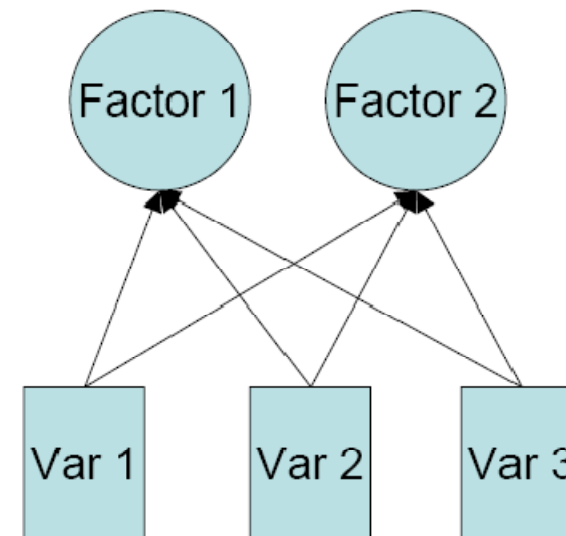


Σχηματική Αναπαράσταση

- Common FA



- PCA



Διαφορές μεταξύ παραγόντων και συνιστωσών



- Οι **παράγοντες** είναι πραγματικές **λανθάνουσες μεταβλητές**, οι οποίες προκαλούν τη συνδιακύμανση μεταξύ των μεταβλητών
- Οι **συνιστώσες** είναι **εμπειρικά καθορισμένα αθροίσματα μεταβλητών**, χωρίς απαραίτητα να υπάρχει θεωρητική τεκμηρίωση της εμφάνισής τους.

Η μέθοδος «εξαγωγής» των παραγόντων



- Αρχικά «εξάγεται» **ο πρώτος παράγοντας ή συνιστώσα**, ο οποίος ερμηνεύει το μεγαλύτερο δυνατό **ποσοστό της διακύμανσης** ανάμεσα στα στοιχεία (items) και τον παράγοντα (συσχέτιση)
- Στη συνέχεια «εξάγεται» **ο επόμενος παράγοντας ή συνιστώσα**, ο οποίος ερμηνεύει το μεγαλύτερο δυνατό ποσοστό της διακύμανσης που έχει απομείνει από την ερμηνεία του πρώτου παράγοντα.
- Στη συνέχεια «εξάγεται» **ο επόμενος παράγοντας ή συνιστώσα** μέχρι να μην μείνει ποσοστό διακύμανσης που δεν ερμηνεύεται από τα στοιχεία που μελετάμε

Πόσοι παράγοντες θα πρέπει τελικά να εξαχθούν;



- Βασίζεται στις **ιδιοτιμές** (eigenvalues)
 - Κρατάμε όσους παράγοντες έχουν ιδιοτιμή πάνω από 1 (Guttman – Kaiser)
 - Χρήση του Scree test (Cattell)
 - Κρατάμε τους παράγοντες που ερμηνεύουν το 70-80% της συνολικής διακύμανσης

Κριτήριο Guttman - Kaiser (Ιδιοτιμές > 1)



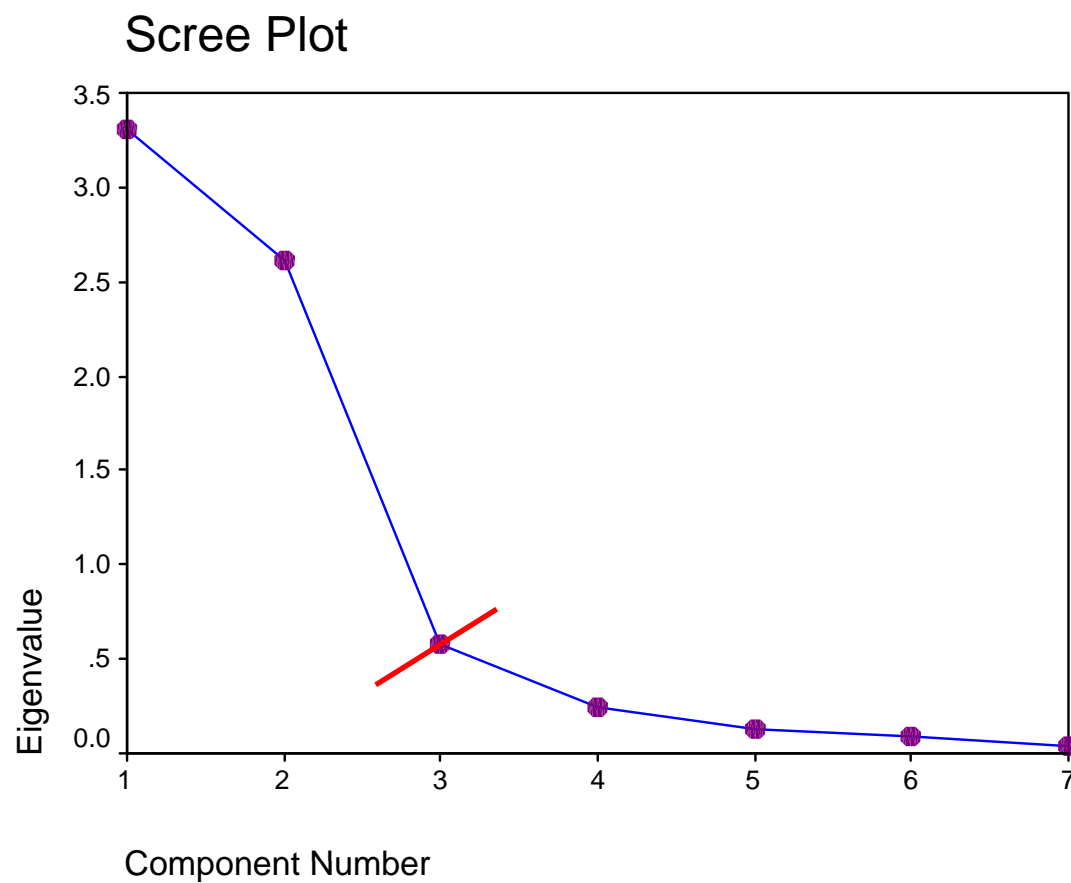
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9,169	32,748	32,748	9,169	32,748	32,748	3,889	13,888	13,888
2	1,934	6,907	39,655	1,934	6,907	39,655	3,474	12,409	26,297
3	1,739	6,212	45,867	1,739	6,212	45,867	2,848	10,170	36,467
4	1,338	4,778	50,644	1,338	4,778	50,644	2,433	8,688	45,154
5	1,173	4,190	54,835	1,173	4,190	54,835	2,197	7,846	53,000
6	1,039	3,710	58,545	1,039	3,710	58,545	1,553	5,545	58,545
7	,983	3,510	62,055						
8	,876	3,128	65,183						
9	,813	2,902	68,085						
10	,769	2,746	70,831						
11	,747	2,669	73,500						
12	,679	2,426	75,927						
13	,634	2,266	78,192						
14	,599	2,140	80,332						
15	,572	2,044	82,376						
16	,545	1,945	84,321						
17	,506	1,806	86,127						
18	,491	1,752	87,879						
19	,452	1,615	89,494						
20	,442	1,577	91,071						
21	,398	1,422	92,493						
22	,393	1,405	93,898						
23	,368	1,316	95,214						
24	,314	1,123	96,337						
25	,305	1,091	97,428						
26	,264	,942	98,371						
27	,237	,847	99,218						
28	,219	,782	100,000						

Extraction Method: Principal Component Analysis.



Κριτήριο Scree Test



Η περιστροφή των παραγόντων



- Η περιστροφή των παραγόντων είναι απαραίτητη για την **ευκολότερη ερμηνεία των παραγόντων** (ή συνιστωσών) που έχουν προκύψει από την ανάλυση.
 - **Ορθογώνια περιστροφή** (orthogonal)
 - **Πλάγια περιστροφή** (oblique)

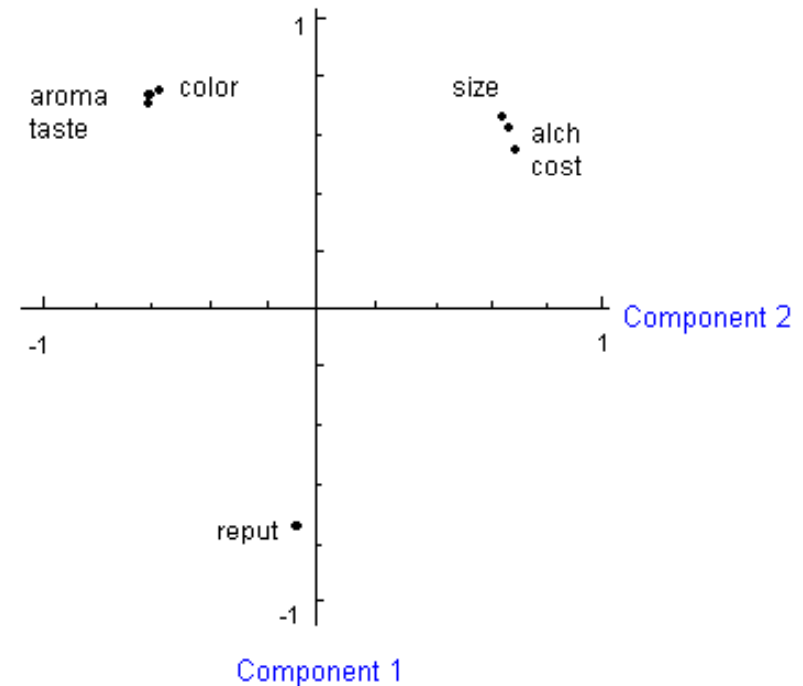


Παράδειγμα: Χωρίς περιστροφή

Component Matrix

	Component	
	1	2
COLOR	.760	-.576
AROMA	.736	-.614
REPUTA	-.735	-.071
TASTE	.710	-.646
COST	.550	.734
ALCOHC	.632	.699
SIZE	.667	.675

Extraction Method: Principal Compone
a.2 components extracted.





Παράδειγμα: Με περιστροφή

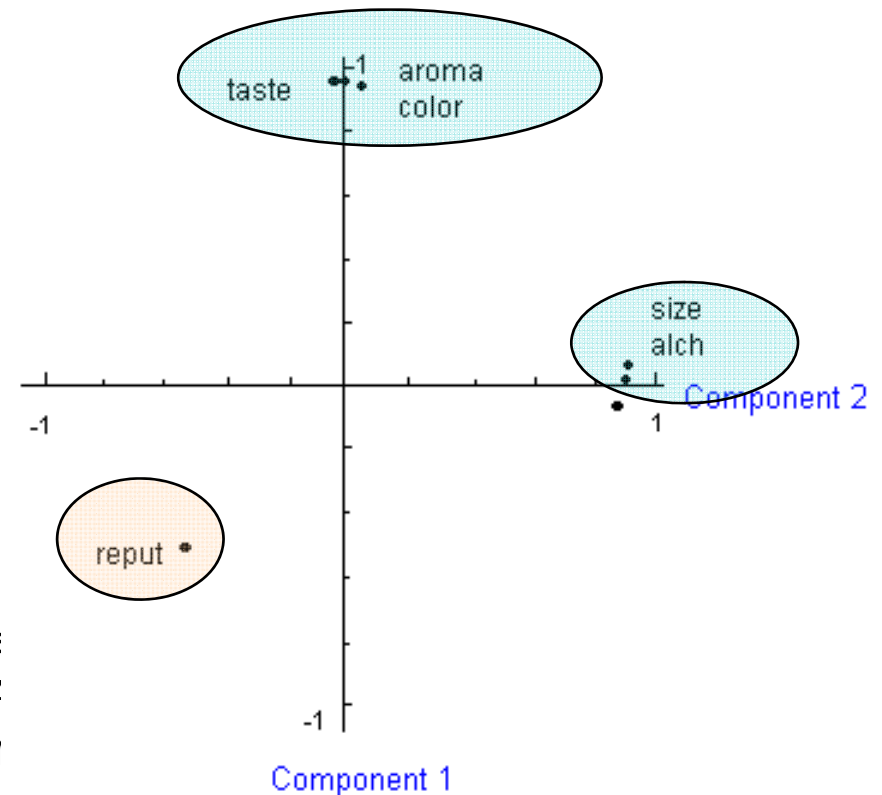
Rotated Component Matrix

	Component	
	1	2
TASTE	.960	-.028
AROMA	.958	1.E-02
COLOR	.952	6.E-02
SIZE	7.E-02	.947
ALCOHOL	2.E-02	.942
COST	-.061	.916
REPUTAT	-.512	-.533

Extraction Method: Principal Component Analy

Rotation Method: Varimax with Kaiser Normaliz

a. Rotation converged in 3 iteratio





Ορθογώνια Περιστροφή

- Την εφαρμόζουμε ότι γνωρίζουμε ή υποθέτουμε με βάση τη θεωρία ότι οι παράγοντες που θα προκύψουν θα είναι **ανεξάρτητοι** μεταξύ τους

- **Varimax**
- Quartimax
- Equamax



Πλάγια περιστροφή

- Την εφαρμόζουμε ότι γνωρίζουμε ή υποθέτουμε με βάση τη θεωρία ότι οι παράγοντες που θα προκύψουν **συσχετίζονται** μεταξύ τους

- **Direct Oblimin**
- Promax
- Quartimin



Φορτίσεις (Factor Loadings)

- Αποτελούν το **βάρος** του της κάθε μεταβλητής (στοιχείου) στον κάθε παράγοντα (δείκτη συσχέτισης)
- Αποτελούν βασικές πληροφορίες για την **ερμηνεία** των παραγόντων
- Όσο **υψηλότερη** η φόρτιση τόσο πιο εύκολη η ερμηνεία του παράγοντα
- Μια φόρτιση κρίνεται σημαντική όταν είναι πάνω από **0.30**
- Σημαντικό ρόλο παίζουν και οι **δευτερογενείς φορτίσεις** (υψηλές φορτίσεις σε δύο παράγοντες ταυτόχρονα)

Rotated Component Matrix^a

	Component					
	1	2	3	4	5	6
ghq25	,824	-,002	,157	,050	-,011	,142
ghq27	,779	,131	,194	,034	,137	,162
ghq28	,768	,206	,120	,011	,076	,203
ghq24	,722	,244	,089	,169	,220	,126
ghq23	,588	,362	,146	,309	,214	,003
ghq22	,531	,340	,227	,350	,188	-,092
ghq26	,463	,283	,264	,317	,140	-,143
ghq18	,127	,734	,246	-,040	,062	,183
ghq19	,287	,705	-,054	,171	,135	-,025
ghq20	,203	,686	,129	,024	,228	,035
ghq17	,128	,685	,261	,133	-,025	,127
ghq15	,104	,577	,012	,284	,050	-,052
ghq16	,020	,535	,110	-,022	,159	,280
ghq2	,154	,059	,681	,150	,188	-,217
ghq1	,153	,209	,658	,011	,126	,066
ghq3	,187	,055	,657	,180	,244	,177
ghq21	,120	,319	,534	,014	-,026	,261
ghq10	,116	,051	,470	,370	,100	,173
ghq12	,179	,141	-,004	,759	,147	,042
ghq14	,087	,130	,369	,612	,136	,146
ghq11	,030	,027	,174	,608	,010	,409
ghq13	,176	,170	,462	,465	,152	,126
ghq5	,102	,175	,174	,062	,854	,055
ghq6	,202	,170	,235	,167	,768	,077
ghq7	,142	,192	,104	,146	,440	,438
ghq4	,297	,032	,350	,230	,426	,135
ghq8	,273	,094	,156	,205	,204	,656
ghq9	,259	,330	,042	,274	-,015	,501

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.



Communalities

Το ποσοστό της διακύμανσης που εξηγείται από τους παράγοντες

	Initial	Extraction
ghq1	1,000	,520
ghq2	1,000	,596
ghq3	1,000	,593
ghq4	1,000	,464
ghq5	1,000	,807
ghq6	1,000	,749
ghq7	1,000	,475
ghq8	1,000	,622
ghq9	1,000	,504
ghq10	1,000	,414
ghq11	1,000	,568
ghq12	1,000	,652
ghq13	1,000	,529
ghq14	1,000	,575
ghq15	1,000	,429
ghq16	1,000	,403
ghq17	1,000	,587
ghq18	1,000	,654
ghq19	1,000	,630
ghq20	1,000	,582
ghq21	1,000	,470
ghq22	1,000	,615
ghq23	1,000	,639
ghq24	1,000	,681
ghq25	1,000	,727
ghq26	1,000	,504
ghq27	1,000	,708
ghq28	1,000	,693