

ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΑΝΤΙΣΤΟΙΧΙΩΝ

- Η Παραγοντική Ανάλυση των Αντιστοιχιών (Analyse Factorielle des Corespondances -AFC-) είναι μία μέθοδος με την οποία μπορούμε να μελετήσουμε τη συμπεριφορά και τις αλληλεπιδράσεις μεταξύ ποιοτικών ή ποσοτικών μεταβλητών χωρισμένων σε κλάσεις, καθώς και κάθε εξάρτηση μεταξύ «**γραμμών**» και «**στηλών**» του πίνακα δεδομένων.
- Η μαθηματική διατύπωση της ανάλυσης οφείλεται στον Γάλλο καθηγητή του Πανεπιστημίου Pierre et Marie CURIE PARIS VI **JEAN PAUL BENZECRI**
- Κατάλληλοι πίνακες για την εφαρμογή της μεθόδου είναι οι πίνακες συμπτώσεων απολύτων συχνοτήτων δύο ποιοτικών μεταβλητών X και Y, οι πίνακες αξιολόγησης- κατάταξης, οι διαχρονικοί πίνακες συμπτώσεων, οι διαζευκτικοί πίνακες και οι πίνακες Burt.
- Ένα σημαντικό πλεονέκτημα της μεθόδου είναι η εξέταση ενός φαινομένου χρησιμοποιώντας περισσότερες από δύο μεταβλητές, αντιμετωπίζοντας έτσι ένα μεγάλο μειονέκτημα πολλών μεθόδων της κλασικής Στατιστικής. Για τον λόγο αυτό χρειάστηκε να χρησιμοποιηθούν τα μαθηματικά του Ευκλείδειου n -διάστατου χώρου, με αποτέλεσμα:
 - 1ον)** κάθε φαινόμενο επειδή από την φύση του είναι σύνθετο στο οποίο υπεισέρχονται πληθώρα παραγόντων, να εξετάζεται με βάση την συνολική αλληλεξάρτηση των παραγόντων αυτών και όχι πλέον δύο-δύο χωριστά.
 - 2ον)** την διανυσματοποίηση ποιοτικών χαρακτηριστικών, δίχως την ανάγκη να χρησιμοποιηθούν οι λεγόμενες ψευδομεταβλητές, που από την ονομασία τους και μόνο καταδεικνύεται η λογική αφαίρεση που πραγματοποιείται από τους θιασώτες της κλασικής Στατιστικής (κυρίως από τους Οικονομέτες).

Η γενική μορφή ενός πίνακα συμπτώσεων είναι η παρακάτω:

Πίνακας 2.1

		Μεταβλητή Y		Περιθωριακή στήλη
		j	j	
Μεταβλητή X	i	k(i,j)	k(i)	k(i) άθροισμα i γραμμής
	k(j)	k(j)	k	k Σύνολο
Περιθωριακή γραμμή				

όπου

$$k_i = \sum_{j=1}^p a_{ij} \quad k_{ij} = k(i)$$

$$k_j = \sum_{i=1}^n a_{ij} \quad k_{ij} = k(j)$$

$$k = \sum_{i=1}^n \sum_{j=1}^p a_{ij} \quad k_{ij}$$

Στην αρχή της ανάλυσης τροποποιούμε τον πίνακα 2.1 σ' ένα πίνακα **σχετικών συχνοτήτων** (πίνακας 2.2), δηλαδή αντικαθίστανται τα k_{ij} με τις σχετικές συχνοτήτες f_{ij} , οι οποίες υπολογίζονται από τη σχέση 2.4.

$$f_{ij} = \frac{k_{ij}}{k} \quad 2.4$$

Πίνακας 2.2
Πίνακας σχετικών συχνοτήτων

A \ B	b1	b2	b _j	b _p	
a1						
a2						
.						
a _i			 f _{ij}		f _i
.						
a _n						
				f _j		

όπου $f_i = \frac{k_i}{k}$ και $f_j = \frac{k_j}{k}$

Μάζα ενός στοιχείου του συνόλου I (ή του J)

Σε κάθε στοιχείο του συνόλου $I=\{1,\dots,n\}$ προσαρτάται η ποσότητα (2.9) η οποία καλείται μάζα του «ατόμου» i .

$$f_i = \frac{k(i)}{k} \quad (2.9)$$

Ενώ

σε κάθε στοιχείο του συνόλου $J=\{1,\dots,p\}$ προσαρτάται μάζα ίση με

$$f_j = \frac{k(j)}{k} \quad (2.10)$$

Η μάζα κάθε στοιχείου i ή j μετρά την σχετική σπουδαιότητα του στοιχείου

Ο πίνακας των "προφίλ" των γραμμών

Για να είναι συγκρίσιμες δύο γραμμές i και i' του πίνακα 2.2 των σχετικών συχνοτήτων, δημιουργούμε τον πίνακα των προφίλ των γραμμών, (οι οποίες αποτελούν τις ποσοστιαίες κατανομές στο εσωτερικό κάθε γραμμής χωριστά, ή αλλιώς τις σχετικές κατανομές τους), χρησιμοποιώντας τη σχέση 2.11.

$$f_j^i = \frac{f_{ij}}{f_i} \quad (2.11)$$

ή

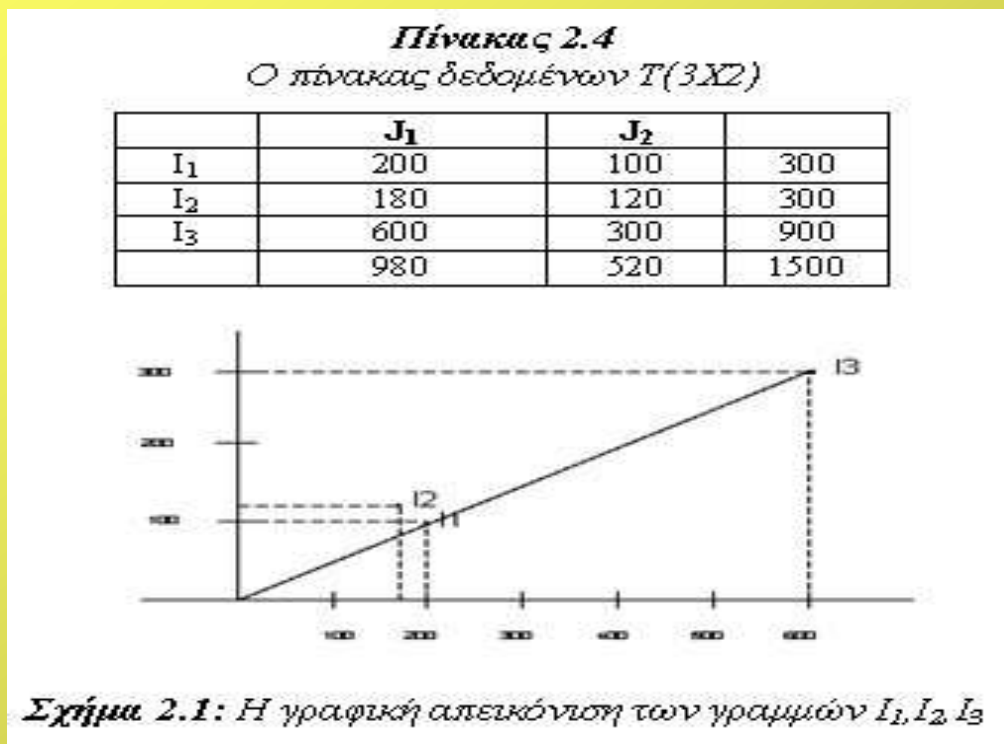
$$f_j^i = \frac{k_{ij}}{k_i}$$

Ο όρος f_j^i αποτελεί την αναλογία της διαβάθμισης j που παρουσιάζεται στην i -οστή γραμμή, ενώ το προφίλ $f_j^i = \{f_j^i \mid j \in J\}$ είναι το σύνολο των αναλογιών των διαβαθμίσεων $J = \{j=1, \dots, p\}$ που παρουσιάζονται στην i -οστή γραμμή.

Πίνακας 2.3
Πίνακας f_j^i των προφίλ των γραμμών

Κλάσεις	$b_1 \dots \dots \dots b_i \dots \dots \dots b_p$	
a_1	.	
.	.	
a_i $f_i^i = f_{ii}/f_i$	1
.	.	
a_n	.	

Έστω ένας πίνακας $T(3 \times 2)$. Η απεικόνιση των γραμμών του πίνακα στο επίπεδο παρουσιάζεται στο σχήμα 2.1.

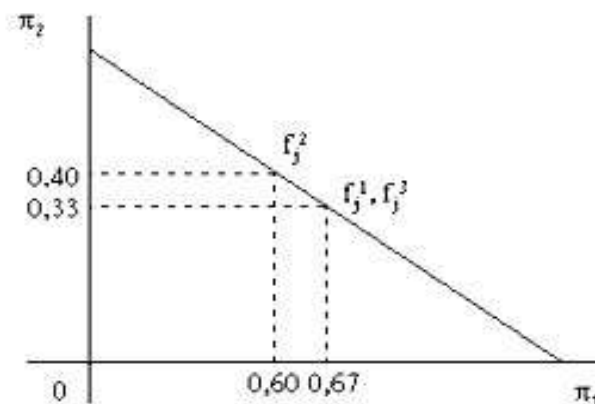


Προσδιορίζουμε τα προφίλ των γραμμών I_1, I_2, I_3

$$f_j^1 = (200/300, 100/300) = (0.67, 0.33)$$

$$f_j^2 = (180/300, 120/300) = (0.60, 0.40)$$

$$f_j^3 = (600/900, 300/900) = (0.67, 0.33)$$



Σχήμα 2.2: Γραφική απεικόνιση των προφίλ των γραμμών

Ο πίνακας των "προφίλ" των στηλών

$$f_i^j = \frac{k_{ij}}{k_j}$$

Πίνακας 2.5

Πίνακας των προφίλ των στηλών

κλάσεις	b₁.....b_j.....b_p
a ₁	· · · · ·
·	· · · · ·
a _i	f _i ^j = f _{ij} /f _j
·	· · · · ·
a _n	· · · · ·
	1

Ο δείκτης έλξης-άπωσης

Ορίζουμε ως δείκτη έλξης-άπωσης μεταξύ μιας γραμμής i και μιας στήλης j την ποσότητα που ορίζεται από τη σχέση 2.18.

$$d_{ij} = \frac{f_{ij}}{f_i \cdot f_j} \quad (2.18)$$

Όταν λοιπόν έχουμε

$d_{ij} > 1$	υπάρχει έλξη	μεταξύ των i και j
$d_{ij} < 1$	»	άπωση » »
$d_{ij} = 1$	δεν προκύπτει κανένα συμπέρασμα	

Έλεγχος της ανεξαρτησίας μεταξύ των δύο ποιοτικών μεταβλητών

Χρησιμοποιείται η στατιστική χ^2

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{(e_{ij} - \frac{k_i \times k_j}{k})^2}{\frac{k_i \times k_j}{k}} \quad (2.19)$$

Την τιμή του χ^2 που προκύπτει από τη σχέση 2.19 τη συγκρίνουμε με την κριτική τιμή $\chi^2_{(n-1)(p-1),0.05}$

οπότε αν βρεθεί ότι ισχύει

$$\chi^2 < \chi^2_{(n-1)(p-1),0.05}$$

θεωρούμε ότι ο δεσμός μεταξύ των δύο μεταβλητών δεν είναι στατιστικά σημαντικός.

Συμβολή του φατνίου στο χ^2

Η διαφορά μεταξύ της Π.Α.Α και του τεστ του χ^2 είναι ότι, αν δύο φατνία παρουσιάζουν την ίδια απόκλιση από την ανεξαρτησία, η Π.Α.Α υποδεικνύει αν έχουν και την ίδια **σημαντικότητα**, κάτι που δεν προκύπτει από το τεστ του χ^2 .

	M	K	A	ΠΓ
I ₁	13	2	5	20
I ₂	20	2	8	30
I ₃	10	5	5	20
I ₄	7	1	22	30
ΠΣ	50	10	40	100

	M	K	A		M	K	A		M	K	A
I ₁	13	2	5	-	10	2	8	=	3	0	-3
I ₂	20	2	8	-	15	3	12	=	5	-1	-4
I ₃	10	5	5	-	10	2	8	=	0	3	-3
I ₄	7	1	22	-	15	3	12	=	-8	-2	10

Πίνακας 2.6

Φατνίο	Παρατηρούμενες τιμές	Θεωρητικές τιμές	Απόκλιση
(I ₁ ,M)	13	10	3
(I ₃ ,K)	5	2	3

Η ταινία I₁ αντιπροσωπεύει το 20% των θεατών και η απόκλιση των 3 θεατών αποτελεί τα 3/10 της θεωρητικής τιμής. Όσο για την ταινία I₃, που αντιπροσωπεύει και αυτή το 20% των θεατών, η απόκλιση των 3 θεατών αποτελεί τα 3/2 της θεωρητικής τιμής.

Πίνακας 2.7

Φατνίο	Απόκλιση	Απόκλιση/Θεωρητική	Σταθμισμένη απόκλιση
(I ₁ ,M)	3	3/10=0,3	3x0,3=0,9
(I ₃ ,K)	3	3/2=1,5	3x1,5=4,5

Η σταθμισμένη απόκλιση ονομάζεται γενικώς **συμβολή του φατνίου στο χ^2** .

Το νέφος $N(I)_J$ των προφίλ των γραμμών

Το σύνολο των προφίλ γραμμών με τα αντίστοιχα βάρη τους αποτελεί το νέφος $N(I)_J$, του οποίου το κέντρο βάρους ορίζεται από τη σχέση 2.21.

$$G_j = \mathring{a} \sum_{i=1}^n f_i \times f_J^i \quad (2.21)$$

Αυτό σημαίνει πως η j -οστή συντεταγμένη του κέντρου βάρους ισούται με

$$\mathring{a} \sum_{i=1}^n f_i \times f_J^i = \mathring{a} \sum_{i=1}^n f_i \frac{f_{ij}}{f_i} = \mathring{a} \sum_{i=1}^n f_{ij} = f_j$$

Άρα το διάνυσμα $G_J = \{f_J | J=1, \dots, p\}$ έχει ως συντεταγμένες τα στοιχεία της περιθωριακής γραμμής του πίνακα των σχετικών συχνοτήτων.

Η μετρική του χ^2

Η απόσταση $d^2(f_j^i, f_j^{i'})$ μεταξύ δύο προφίλ γραμμών ορίζεται από τη σχέση 2.24.

$$d^2(f_j^i, f_j^{i'}) = \sum_{j=1}^p \frac{1}{f_j} (f_j^i, f_j^{i'})^2 \quad (2.24)$$

η οποία ονομάζεται **απόσταση χ^2** και θεσπίστηκε από τον Γάλλο καθηγητή **J.P Benzecri**

Η αδράνεια του νέφους $N(I)_J$ των προφίλ των γραμμών

Η αδράνεια του νέφους $N(I)_J$ των προφίλ των γραμμών υπολογίζεται από τον παρακάτω τύπο

$$I(N, f_J) = \frac{x^2}{k} = \varphi^2$$

Πρέπει να σημειωθεί ότι όσο η αδράνεια είναι μεγαλύτερη, τόσο μεγαλύτερη είναι η αλληλεξάρτηση μεταξύ των «γραμμών» και των «στηλών», ενώ η μέγιστη τιμή που μπορεί να πάρει είναι

$$\min\{\text{card}(I), \text{card}J\}-1.$$

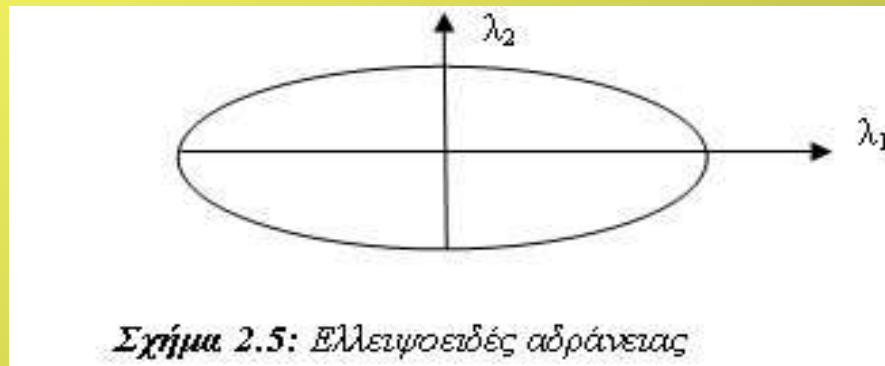
Η συνολική αδράνεια αποτελεί **δείκτη της διασποράς** των σημείων του νέφους. Έτσι η σύγκριση των αδρανειών που αντιστοιχούν σε κάθε άξονα, πληροφορεί για την μορφή του νέφους των σημείων. Αν λοιπόν οι πρώτες χαρακτηριστικές ρίζες είναι η μία κοντά στην άλλη τότε η διασπορά των σημείων είναι ομοιογενής, δηλαδή δεν υπάρχει κάποια **προνομιούχος διεύθυνση** και το νέφος παρουσιάζεται σχεδόν **σφαιρικό**.

Εάν όμως οι χαρακτηριστικές ρίζες λ_i είναι σαφώς διαφορετικές, τότε τα σημεία αναπτύσσονται κατά μήκος περισσότερων διευθύνσεων, που σημαίνει ύπαρξη συγκεκριμένων τάσεων.

Παραγοντικοί άξονες στο επίπεδο

Η αδράνεια του 1^{ου} παραγοντικού άξονα προσδιορίζεται από την χαρακτηριστική τιμή λ_1 , ενώ του 2^{ου} παραγοντικού άξονα από την δεύτερη χαρακτηριστική ρίζα λ_2 κ.ο.κ.

Επομένως μεταξύ των χαρακτηριστικών ριζών υφίσταται η σχέση $\lambda_1 > \lambda_2 > \dots > \lambda_n$, όπου λ_n η χαρακτηριστική ρίζα του n -ιοστού παραγοντικού άξονα στην περίπτωση που έχουμε ένα πολυδιάστατο χώρο R^n του οποίου το νέφος θα είναι ένα ελλειψοειδές με n άξονες.



Η φθίνουσα εξέλιξη των τιμών λ_i είναι όλο και πιο έντονη, όσο απομακρύνονται οι «γραμμές» και οι «στήλες» ενός πίνακα δεδομένων από την κατάσταση ανεξαρτησίας

Η μορφολογία του νέφους και η κατανομή των σημείων του νέφους σε καθένα από τους άξονες αδράνειας Δ_α επιτρέπουν την «ανάγνωση» της πληροφορίας που περιέχεται στον πίνακα δεδομένων.

Η ανάλυση του νέφους $N(I)_J$ των προφίλ των γραμμών

Οι προβολές των σημείων f_j^i του νέφους $N(I)_J$ πάνω στους παραγοντικούς άξονες Δ_a ($a=1, \dots, p-1$), οι οποίες συμβολίζονται με $F_a(i)$ (όπου i = τυχόν προφίλ γραμμή), αποτελούν τις συντεταγμένες των σημείων αυτών πάνω στους άξονες Δ_a .

Κάθε συντεταγμένη $F_a(i)$ σχετική με τον παραγοντικό άξονα a ονομάζεται **παράγοντας a του προφίλ i**

Σε κάθε παραγοντικό άξονα Δ_a ισχύουν οι παρακάτω σχέσεις

$$\overset{\circ}{\mathbf{a}} \sum_{i=1}^n f_i \times F_a(i) = 0 \quad \text{και} \quad \overset{\circ}{\mathbf{a}} \sum_{i=1}^n f_i \times (F_a(i))^2 = \lambda_a$$

ενώ για δύο διαφορετικούς παραγοντικούς άξονες Δ_r και Δ_s ισχύει:

$$\overset{\circ}{\mathbf{a}} \sum_{i=1}^n f_i \times F_r(i) F_s(i) = 0$$

Ποιότητα προβολής των προφίλ των γραμμών

$$COR_a(i) = \frac{F_a^2(i)}{d^2(i, G)} \quad \text{όπου} \quad d^2(i, G) = \overset{\circ}{\mathbf{a}} \sum_{a=1}^p F_a^2(i) \quad \text{και} \quad \overset{\circ}{\mathbf{a}} \sum_{a=1}^{p-1} COR_a(i) = 1$$

Η συμβολή κάθε στατιστικής μονάδας στη διαμόρφωση του παραγοντικού άξονα Δ_α

$$CTR_\alpha(i) = \frac{f_i \cdot F_\alpha^2(i)}{\lambda_\alpha}$$

Η ανάλυση του νέφους $N(J)_i$ των προφίλ των στηλών

Η απόσταση $d^2(f_j^i, f_{j'}^i)$ μεταξύ δύο προφίλ στηλών ορίζεται από τη σχέση 2.50.

$$d^2(f_j^I, f_{j'}^I) = \mathring{\mathbf{a}} \sum_{i=1}^p \frac{1}{f_i} (f_j^I, f_{j'}^I)^2 \quad (2.50)$$

Αποδεικνύεται [Benzecri J.P., 1980,σελ.80] πως η ανάλυση του νέφους $N(J)_i$ δίνει τις ίδιες χαρακτηριστικές ρίζες με εκείνες του νέφους $N(I)_j$. Οπότε οι συντεταγμένες των μεταβλητών στους παραγοντικούς άξονες δίνονται από την παρακάτω σχέση:

$$G_a(j) = \sqrt{\lambda_a} \cdot \varphi_a^j$$

όπου φ_a^j το χαρακτηριστικό διάνυσμα του παραγοντικού άξονα Δ_a

και

λ_a η χαρακτηριστική ρίζα που αντιστοιχεί στον παραγοντικό άξονα Δ_a

Δυϊκή ανάλυση

➤ Η **δυϊκή ανάλυση** μεταξύ των νεφών $N(I)_J$ και $N(J)_I$, οδηγεί όπως είναι φυσικό να υπάρχει στενή σχέση μεταξύ των παραγόντων F_a και G_a . Η σχέση αυτή καθορίζεται βάσει των παρακάτω τύπων:

$$F_a(i) = \frac{1}{\sqrt{\lambda_\alpha}} \mathring{\mathbf{a}}_{j=1}^p \frac{f_{ij}}{f_i} \times G_a(j) \quad (2.55)$$

$$G_a(j) = \frac{1}{\sqrt{\lambda_\alpha}} \mathring{\mathbf{a}}_{i=1}^n \frac{f_{ij}}{f_j} \times F_a(i) \quad (2.56)$$

➤ Οι σχέσεις 2.55 και 2.56 αποτελούν τους τύπους «**μετάβασης**» μεταξύ των παραγόντων των δύο νεφών $N(I)_J$ και $N(J)_I$. Οπότε

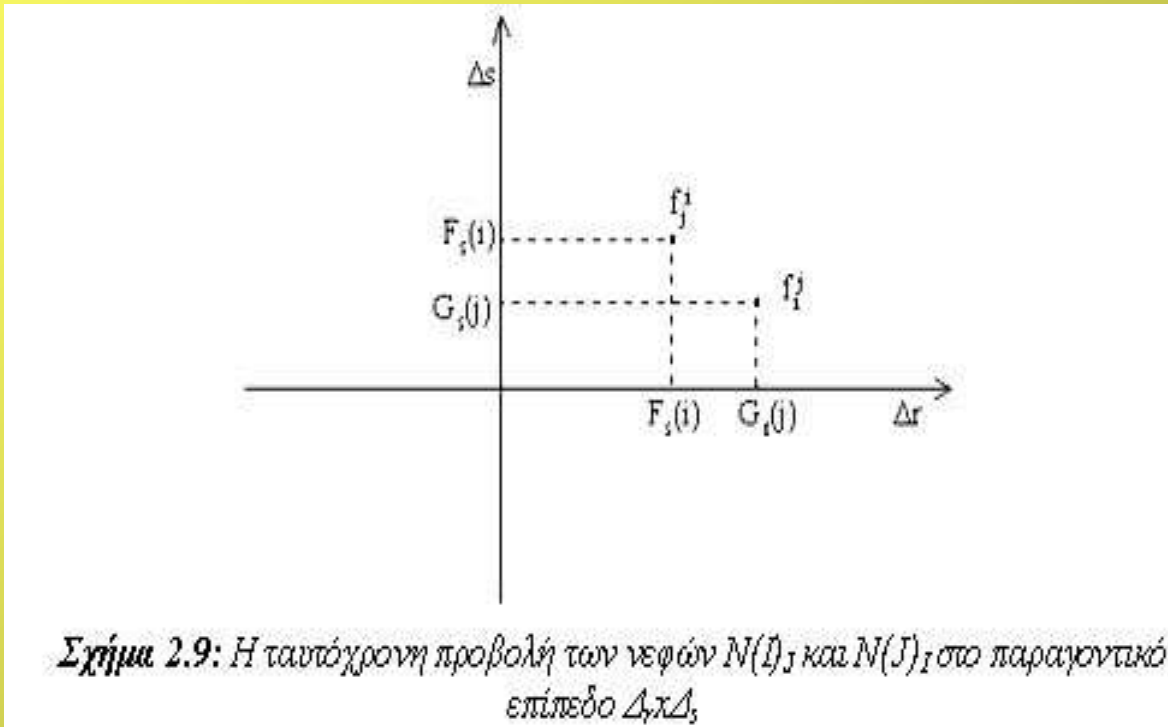
$$G_a(j) \times \sqrt{\lambda_\alpha} = \mathring{\mathbf{a}}_{i=1}^n \frac{f_{ij}}{f_i} \times F_a(i)$$

➤ Η χαρακτηριστική αυτή ιδιότητα υπήρξε βασικά ο λόγος για τον οποίο ονομάστηκε η συγκεκριμένη ανάλυση **Παραγοντική Ανάλυση των Αντιστοιχιών**

➤ Η ιδιότητα αυτή αποτελεί την **βαρυκεντρική αρχή** που διέπει την ανάλυση των νεφών των προφίλ των γραμμών και των στηλών

Τα παραγοντικά επίπεδα

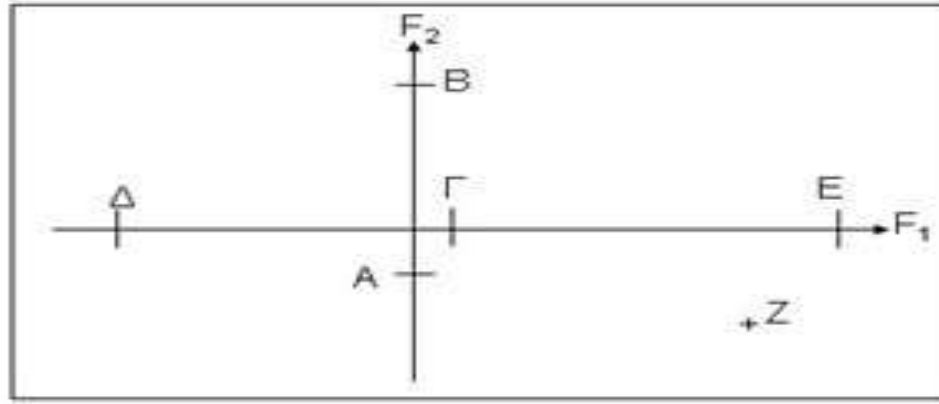
Επειδή με την ανάλυση των νεφών $N(I)_J$ και $N(J)_I$ προκύπτουν οι ίδιες χαρακτηριστικές ρίζες λ_a ($a=1, \dots, p-1$) και για τα δύο νέφη, είναι εφικτή η γραφική απεικόνιση των δύο νεφών στο ίδιο σύστημα αξόνων $\Delta_r \times \Delta_s$.



Πρέπει πάντως να προσέξουμε ότι με την Παραγοντική Ανάλυση των Αντιστοιχιών παρουσιάζουμε ανάγλυφη την δομή των αποκλίσεων από την ανεξαρτησία καθώς και τους λόγους κάθε απόκλισης όχι όμως την έντασή τους.

Βοήθειες για την ερμηνεία ενός παραγοντικού άξονα

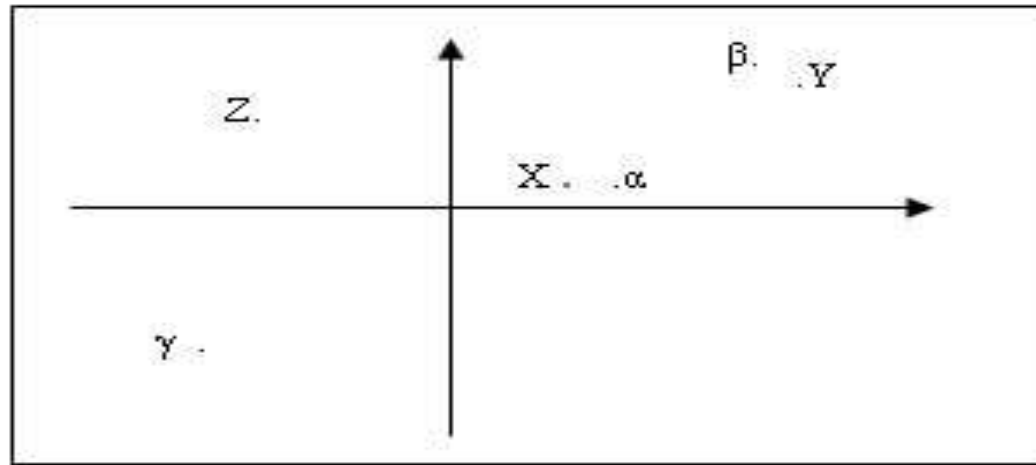
Τα σημεία A,B,Γ,Δ,E,Z θεωρούνται προφίλ γραμμές



Σχήμα 2.10: Παραγοντικό επίπεδο 6 προφίλ γραμμών

- ✓ Τα σημεία **A** και **B** είναι σημεία που συμπεριφέρονται όπως η μέση κατάσταση, η οποία αντιστοιχεί στο προφίλ της περιθωριακής γραμμής, ενώ τα σημεία **Δ**, **E**, **Z** είναι σημεία ακραίας συμπεριφοράς, με τα σημεία **E** και **Z** να αντιπαράθενται με το σημείο **Δ**.
- ✓ Η **αντιπαράθεση** αυτή μπορεί να προσδώσει στον παραγοντικό άξονα κάποια ιδιότητα που θα τον χαρακτηρίζει. Η ιδιότητα αυτή αποτελεί αυτό που ονομάζεται **υποβόσκουσα τάση**.
- ✓ Για τα σημεία **Δ** και **E** η ποιότητα προβολής (COR) θα είναι ίση με 1, ενώ η συνεισφορά στην ερμηνευόμενη αδράνεια από τον παραγοντικό άξονα (CTR) θα είναι υψηλή.
- ✓ Τα σημεία **A** και **B** τα οποία είναι τοποθετημένα σε μία διεύθυνση κάθετη προς τον 1ο παραγοντικό άξονα έχουν ποιότητα προβολής ίση με 0. Αυτό σημαίνει πως δεν συμμετέχουν καθόλου στην διαμόρφωση του 1ου παραγοντικού άξονα, καθώς έχουν συντεταγμένες ίσες με 0, άρα και CTR=0.
- ✓ Το σημείο **Z** έχει μία μεσαία προβολή. Επομένως παρότι χαρακτηρίζεται από τον 1ο παραγοντικό άξονα, η απόκλισή του από τον μέση κατάσταση εξηγείται μερικώς, καθώς συμμετέχει στην ερμηνεία και ο 2ος παραγοντικός άξονας.
- ✓ Τέλος το σημείο **Γ** συμπεριφέρεται περίπου όπως η μέση κατάσταση η οποία αντιστοιχεί στην περιθωριακή κατανομή των γραμμών του πίνακα δεδομένων.

Παγίδες με τις δυϊκές σχέσεις



Σχήμα 2.11: Ιδιομορφίες των δυϊκών σχέσεων

Η διαβάθμιση X βρίσκεται στο **βαρύκεντρο** των σημείων β και γ , οπότε η προσέγγιση με το σημείο α είναι τυχαία. Βέβαια δεν συμβαίνει το ίδιο με την προσέγγιση των σημείων γ και β .

Συνεπώς οι προσεγγίσεις των σημείων των δύο νεφών προς το κέντρο του παραγοντικού επιπέδου είναι επισφαλείς, ενώ μόνο οι προσεγγίσεις στην περιφέρεια του επιπέδου μπορούν να δώσουν ασφαλείς διαπιστώσεις.

Πλήθος ερμηνεύσιμων παραγοντικών αξόνων

Στο ερώτημα αυτό δεν υπάρχει ικανοποιητική απάντηση, καθόσον κάθε κλασική διαδικασία ελέγχου απαιτεί την a priori γνώση της κατανομής των στοιχείων του πίνακα δεδομένων κάτι που συνήθως στην πράξη δεν είναι εφικτό.

Εμπειρικά αν η απώλεια από τους k πρώτους παραγοντικούς άξονες δεν είναι μεγαλύτερη του 25%-30% της συνολικής αδράνειας, τότε η αντλούμενη πληροφορία θεωρείται ικανοποιητική.

Βέβαια στην περίπτωση ανάλυσης ενός λογικού πίνακα 0-1 ο καθηγητής **J.P Benzecri** αναφέρει τα εξής:

Συνήθως συμβολίζουμε με Card Q το πλήθος των ερωτημάτων και με Card J το πλήθος των διαβαθμίσεων των Q ερωτημάτων (με Card συμβολίζουμε τον πληθάρημο ενός συνόλου).

Οπότε στη γενική περίπτωση ενός ερωτηματολογίου με Q ερωτήματα, περιοριζόμαστε στα ποσοστά των παραγοντικών αξόνων των οποίων οι χαρακτηριστικές ρίζες τους είναι μεγαλύτερες του $\frac{1}{s}$, όπου s ο πληθάρημος των Q ερωτημάτων.

Ενώ για τον αντίστοιχο πίνακα Burt το κριτήριο είναι $\frac{1}{s^2}$

Γενικές αρχές εφαρμογής της Π.Α.Α

Οι γενικές αρχές που διέπουν την Παραγοντική Ανάλυση των Αντιστοιχιών είναι οι εξής:

- Οι τιμές των φατνίων ενός πίνακα πρέπει να είναι μη αρνητικοί αριθμοί
- Οι περιθωριακές κατανομές ενός πίνακα δεδομένων πρέπει να έχουν φυσική σημασία. Με άλλα λόγια τα φατνία του πίνακα δεδομένων πρέπει να είναι συχνότητες. Αν τα δεδομένα προέρχονται από ετερογενείς μεταβλητές λ.χ τιμές ποσοστών, τιμές μιας ποσοτικής μεταβλητής και διαβαθμίσεις ενός ποιοτικού χαρακτηριστικού, τότε για να αναλυθεί ένας παρόμοιος πίνακας δεδομένων με την Π.Α.Α πρέπει προηγουμένως τα δεδομένα να ομογενοποιηθούν ώστε να δημιουργηθεί ο αντίστοιχος λογικός πίνακας 0-1.
- Ο 1ος παράγοντας είναι μία κατ' αρχή προσέγγιση των δεδομένων, η οποία όμως χρειάζεται διόρθωση. Η διόρθωση επιτυγχάνεται με τον 2ο παράγοντα. Εάν δεν ικανοποιούμαστε από την προσέγγιση που προσφέρουν οι δύο πρώτοι παράγοντες συνεχίζουμε με τους υπόλοιπους που διαθέτουμε.

Έλεγχος ανεξαρτησίας του πίνακα συμπτώσεων

Σκόπιμο είναι πριν από μία ανάλυση ενός πίνακα συμπτώσεων με την Π.Α.Α να διαπιστωθεί αν υφίσταται ή όχι ανεξαρτησία μεταξύ των «γραμμών» και των «στηλών» του.

Αν δεν γίνει αποδεκτή η υπόθεση της ανεξαρτησίας τότε υφίσταται ενδιαφέρον να περιγραφεί το πλέγμα των αλληλεξαρτήσεων μεταξύ των στατιστικών μονάδων και των μεταβλητών που τις χαρακτηρίζουν.

Έτσι η υπόθεση της ανεξαρτησίας απορρίπτεται εφόσον η ποσότητα χ^2 η οποία υπολογίζεται βάσει της σχέσης :

$$\chi^2 = I(N,G).k \quad (2.84)$$

είναι μεγαλύτερη από την ποσότητα $\chi_{v,a}^2$ για ένα συγκεκριμένο επίπεδο σημαντικότητας α .

Η ενδιάμεση κατάσταση για ένα πίνακα δεδομένων $T(n \times p)$ μεταξύ ανεξαρτησίας και πλήρους εξάρτησης υπολογίζεται από το παρακάτω τύπο του Tschuprow

$$T^2 = \frac{\varphi^2}{\sqrt{(n-1) \times (p-1)}}$$

όπου $\varphi^2 = \frac{\chi^2}{k}$ και $0 \leq T^2 \leq 1$

Ο τύπος ανασύστασης των δεδομένων

Διατηρώντας τους r πρώτους παραγοντικούς άξονες, έχουμε τη δυνατότητα να ανασυστήσουμε τα αρχικά δεδομένα με βάση τους παράγοντες $F_a(i)$ και $G_a(j)$ χρησιμοποιώντας τον παρακάτω τύπο.

$$f_{ij} = f_i \times f_j + \sum_{a=1}^r \frac{1}{\sqrt{\lambda_a}} F_a(i) \times G_a(j)$$

Συμπληρωματικά στοιχεία

Για μία συμπληρωματική μεταβλητή

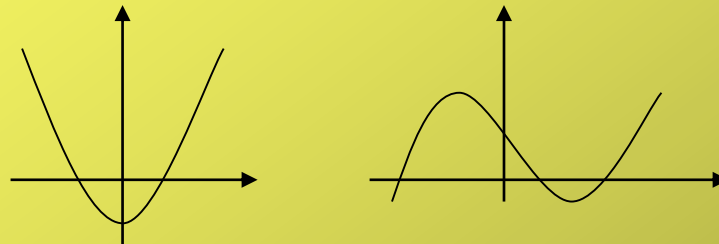
$$G_a(s) = \frac{1}{\sqrt{\lambda_a}} \sum_{i=1}^n \frac{f_{is}}{f_s} F_a(i)$$

Για μία συμπληρωματική στατιστική μονάδα

$$F_a(s) = \frac{1}{\sqrt{\lambda_a}} \sum_{j=1}^p \frac{f_{sj}}{f_s} G_a(j)$$

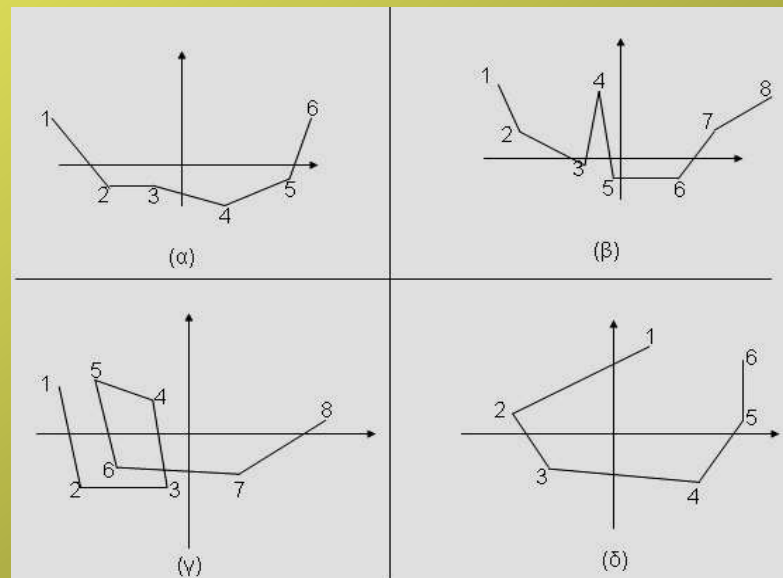
Το φαινόμενο Guttman

Συχνά όταν προβάλλεται το νέφος $N(J)$ στο παραγοντικό επίπεδο εμφανίζονται τα σημεία να σχηματίζουν μία τροχιά όπως στο παρακάτω σχήμα. Η απεικόνιση αυτή υποδηλώνει ότι υφίσταται κάποια σχέση εξάρτησης μεταξύ των μεταβλητών.



Σχήμα 2.13: Τροχιά των ετών όταν παρατηρείται φαινόμενο Guttman

Οι μορφές που μπορούν να εμφανιστούν και να θεωρηθεί ότι συντρέχει το φαινόμενο Guttman παρουσιάζεται στο παρακάτω σχήμα



ΔΙΑΦΟΡΕΤΙΚΗ ΑΝΤΙΛΗΨΗ ΤΗΣ Π.Α.Α ΜΕΤΑΞΥ ΓΑΛΛΙΚΗΣ ΚΑΙ ΑΓΓΛΙΚΗΣ ΣΧΟΛΗΣ

Η διαφορετική αντίληψη περί παραγοντικής ανάλυσης της Αγγλικής Σχολής με εκείνη της Γαλλικής, οφείλεται στη διαφορετική μέθοδο διαχείρισης της διαγωνίου μιας μήτρας διακυμάνσεων-συνδιακυμάνσεων, η οποία ιστορικά αποτελεί την πρώτη μέθοδο παραγοντικής ανάλυσης ενός πολυδιάστατου πίνακα δεδομένων. Στις Αγγλοσαξωνικές χώρες είναι η μόνη μέθοδος στην οποία δίνουν το όνομα Παραγοντική Ανάλυση [Cibois Ph. σελ. 111].

Στη Γαλλία θεωρώντας ότι η Παραγοντική Ανάλυση των Αντιστοιχιών (Π.Α.Α) (Analyse Factorielle des Correspondances) είναι επέκταση της Ανάλυσης σε Κύριες Συνιστώσες (Analyse en Composantes Principales), χρησιμοποιώντας την μετρική του X^2 ο όρος Παραγοντική Ανάλυση περιλαμβάνει το σύνολο των μεθόδων ανάλυσης πολυδιάστατων δεδομένων. Την δε μέθοδο των Αγγλοσαξόνων την ονομάζουν Κλασσική Παραγοντική Ανάλυση (ή ανάλυση παραγόντων των ψυχολόγων).

ΣΥΝΔΕΣΗ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΔΕΣΜΟΥ ΔΥΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΕΝΟΣ ΠΙΝΑΚΑ ΔΙΠΛΗΣ ΕΙΣΟΔΟΥ

Θεωρούμε ότι τα στοιχεία του συνόλου E περιγράφονται βάσει δύο χαρακτηριστικών I και J . Όπως λ.χ. E =χρηματιστηριακή αγορά, I =εισηγμένες εταιρείες και J =κριτήρια αξιολόγησης.

Η πληροφορία τότε η σχετική με το E , που παρέχεται ταυτόχρονα από τα δύο χαρακτηριστικά I και J ισούται με:

$$H(I,J) = \sum_i \sum_j f_{ij} \log \frac{f_{ij}}{f_i \times f_j} \quad \text{Όπου} \quad f_i = \sum_j f_{ij} \quad \text{και} \quad f_j = \sum_i f_{ij}$$

Το $H(I,J)$ μπορεί να θεωρηθεί ως δείκτης εξάρτησης των I και J αν υποθέσουμε ότι το J είναι το εκπεμπόμενο σήμα από μία πηγή πληροφορίας, ενώ το I είναι το λαμβανόμενο σήμα. Τότε ο πίνακας f_{IJ} αποτελεί τον πίνακα σύγχυσης του καναλιού επικοινωνίας. Αν η μεταφορά πληροφορίας είναι άριστη (δίχως θορύβους), ο πίνακας f_{IJ} είναι διαγώνιος (ήτοι $f_{ij}=0$ όταν $i \neq j$), ενώ αν υπεισέρχονται θόρυβοι, τα στοιχεία του πίνακα f_{IJ} εκτός διαγωνίου είναι διάφορα του μηδενός.

Ορίζουμε ως δεσμό των χαρακτηριστικών I και J την ποσότητα $\Delta(I,J)$ η οποία υπολογίζεται βάσει του παρακάτω τύπου

$$\Delta(I,J) = \sum_i \sum_j \frac{(f_{ij} - f_i \times f_j)^2}{f_i \times f_j}$$

Μ' έναν κατάλληλο μετασχηματισμό λαμβάνουμε τη παρακάτω σχέση $H(I,J) \gg \frac{1}{2 \times \text{Log}2} \Delta(I,J)$

Η τελευταία σχέση συνδέει την ποσότητα πληροφορίας $H(I,J)$ με το δεσμό $\Delta(I,J)$, σχέση που υποδηλώνει ότι η εξάρτηση δύο χαρακτηριστικών ισοδυναμεί με έκκληση πληροφορίας.

Πρακτικά σημαίνει ότι η χρήση της μίας έννοιας ή της άλλης οδηγεί σε ανάλογα αποτελέσματα. Επειδή όμως ο δεσμός και η αδράνεια του νέφους $N(I)_J$ των προφίλ των γραμμών προκύπτουν από τον ίδιο τύπο, υποδηλώνει ότι η αδράνεια του νέφους (όπως την ορίζει η Π.Α.Α), αποτελεί μέτρηση της ατλούμενης πληροφορίας από τον πίνακα δεδομένων $T(n \times p)$.