

ΠΑΡΟΥΣΙΑΣΗ

ΤΟΥ

του project

M.A.D

Méthodes d'Analyse des Données

από τον

Δρ. Δημήτριο Νικολάου Καραπιστόλη

ΘΕΣΣΑΛΟΝΙΚΗ 2025

ΠΑΡΟΥΣΙΑΣΗ
ΤΗΣ
ΠΟΛΥΔΙΑΣΤΑΤΗΣ
ΣΤΑΤΙΣΤΙΚΗΣ
ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ

DATA ANALYSIS

✓ Στην αρχή του 20ου αιώνα οι Ευρωπαίοι ψυχολόγοι έψαχναν με τα τεστ που υπέβαλαν στους αρρώστους σε συνδυασμό με τους βαθμούς που συγκέντρωναν από διάφορες μεταβλητές που χρησιμοποιούσαν (μνήμη, ευφυΐα, κ.ά), να δημιουργήσουν σύνθετες μεταβλητές οι οποίες όχι μόνο δεν παρατηρούνται απ' ευθείας από τα αρχικά δεδομένα, αλλά θα ερμήνευαν κατά τον καλύτερο τρόπο τη συμπεριφορά των αρρώστων όσων βέβαια παρουσίαζαν τα ίδια συμπτώματα. Τις μεταβλητές αυτές τις ονόμασαν «**παράγοντες**».

Με τον γνωστό όρο DATA ANALYSIS δεν πρέπει να μας διαφεύγει ότι υπάρχουν δύο διαφορετικές Σχολές οι οποίες χρησιμοποιούν διαφορετικές μαθηματικές μεθόδους προσέγγισης των δεδομένων.
Η Γαλλική Σχολή και η Αγγλοσαξονική Σχολή

- ✓ Στις αρχές της δεκαετίας του '70, ο εμπνευστής της Παραγοντικής Ανάλυσης των Αντιστοιχιών, Γάλλος καθηγητής J.P. Benzecri πρωτοπόρος και θεμελιωτής της Σχολής μαζί με τους μαθητές του, όπως οι Γάλλοι Lebart, Roux, Escoffier, Morineau, Fenelon συντέλεσαν όχι μόνο στη διάδοση αλλά και την καθιέρωση των μεθόδων αυτών στη συνείδηση πολλών ερευνητών σ' ολόκληρο τον κόσμο, δημιουργώντας την **Γαλλική Σχολή της Ανάλυσης Δεδομένων**.
- ✓ Αλλά και στην άλλη μεριά του Ατλαντικού δημιουργήθηκε η λεγόμενη **Αγγλοσαξονική Σχολή** με τους C. Spearman, R.A. Fisher, B. Kruskal, R. S Sheppard, G. Yang κ.λ.π κάτω από το όνομα «**multidimensional scaling**»,

Βασικές Αρχές της Γαλλικής Σχολής

Οι πιο σημαντικές αρχές της Γαλλικής Σχολής της Ανάλυσης Δεδομένων, όπως τις αναφέρει ο ιδρυτής καθηγητής Jean Paul Benzecri είναι οι εξής:

1η αρχή: Η Στατιστική δεν πρέπει να συγχέεται με την θεωρία των **πιθανοτήτων**. Πολλοί στατιστικοί έχουν δομήσει τη μαθηματική στατιστική με πολλές υποθέσεις που σπάνια ή ποτέ δεν ικανοποιούνται στην πράξη.

2η αρχή: Το μοντέλο θα πρέπει να διαμορφώνεται **ακολουθώντας τα δεδομένα** και όχι το αντίστροφο. Αγνοώντας την αρχή αυτή οδηγείται κανείς σ' ένα μεγάλο λάθος της εφαρμογής των μαθηματικών στις επιστήμες της ανθρώπινης συμπεριφοράς. Επιχειρείται δηλαδή η προσαρμογή των δεδομένων στα μοντέλα που έχουν δημιουργηθεί εκ των προτέρων. Ο στόχος αυτός τις περισσότερες φορές δεν υλοποιείται, καθόσον στην ανθρώπινη συμπεριφορά πρωτεύοντα ρόλο διαδραματίζει η μνήμη, οπότε συμπεριφορές του παρελθόντος πολλές φορές δεν είναι παρόμοιες με εκείνες του μέλλοντος, διαδικασία που προκαθορίζεται με την θέσπιση οποιουδήποτε μοντέλου.

3η αρχή: Είναι πρωταρχικής σημασίας να χειρίζεται κανείς πληροφορίες σε όσο το δυνατόν **περισσότερες διαστάσεις**, γεγονός πολύ δύσκολο για την επίλυση ενός φαινομένου με την χρήση των μαθηματικών μοντέλων.

ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΤΗΣ ΓΑΛΛΙΚΗΣ ΣΧΟΛΗΣ

➤ Μεγάλο πλεονέκτημα της νέας σχολής είναι η εξέταση ενός φαινομένου χρησιμοποιώντας **περισσότερες από δύο μεταβλητές**, αντιμετωπίζοντας έτσι ένα μεγάλο μειονέκτημα πολλών μεθόδων της κλασικής Στατιστικής.

➤1ον) Κάθε φαινόμενο επειδή από την φύση του είναι σύνθετο στο οποίο υπεισέρχονται πληθώρα παραγόντων, πρέπει 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 Οικονομέτρους).

Για τον λόγο αυτό χρειάζεται να χρησιμοποιηθούν τα μαθηματικά της Αναλυτικής Γεωμετρίας, δηλαδή του Ευκλείδειου n -διάστατου διανυσματικού χώρου.

ΚΛΑΣΙΚΗ ΣΤΑΤΙΣΤΙΚΗ VS ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

Η μέθοδος των ελαχίστων τετραγώνων, κύρια μέθοδος κατασκευής υποδειγμάτων αποτίμησης κυρίως οικονομικών δεδομένων, θεωρείται από πάρα πολλούς μελετητές ανεπαρκής, επειδή παρουσιάζει αμερόληπτες και συνεπείς εκτιμήσεις των συντελεστών παλινδρόμησης μόνο κάτω από αυστηρές υποθέσεις οι οποίες συνήθως δεν ανταποκρίνονται στη πραγματικότητα.

Όσον αφορά στον κανονικό νόμο παρότι ασχολείται με τη φύση του τυχαίου μάλλον δεν ανταποκρίνεται στην πραγματικότητα

Η ανάγκη λοιπόν να μη θεωρείται εκ των προτέρων ότι ένα φαινόμενο ακολουθεί κάποιο συγκεκριμένο νόμο, οδήγησε στην εφαρμογή νέων στατιστικών μη παραμετρικών μεθόδων, κάτω από την ονομασία **Ανάλυση Δεδομένων** ή όπως αλλιώς την αποκαλώ **Στατιστική δίχως μοντέλα**.

Το βασικό πλεονέκτημα, λοιπόν, του ερευνητή με τις μεθόδους της ανάλυσης δεδομένων ως μη παραμετρικές είναι ότι δεν απαιτείται να λάβει υπόψη του καμιά υπόθεση ως προς τις παραμέτρους που διαμορφώνουν το υπό μελέτη φαινόμενο

Απλά γίνεται προσπάθεια εντοπισμού των αλληλοεξαρτήσεων μεταξύ των παραγόντων (ποσοτικών και ποιοτικών ταυτόχρονα) που ελήφθησαν υπόψη, ώστε να ανακαλυφθεί η συμπεριφορά των στοιχείων που συνθέτουν το υπό μελέτη φαινόμενο και να προσδιοριστεί η δομή του συστήματος που αντιστοιχεί στο φαινόμενο.

ΣΛΟΓΚΑΝ ΤΗΣ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ

Ο ερευνητής για να χρησιμοποιήσει μία παραγοντική μέθοδο, οφείλει να θέσει τα δεδομένα σε ένα πίνακα διαστάσεων $T(I,J)$ όπου

$I=(1,\dots,n)$ το πλήθος των γραμμών (ατόμων ή αντικειμένων)

$J=(1,\dots,p)$ το πλήθος των στηλών (μεταβλητών)

Η Παραγοντική Ανάλυση θέτει σε γεωμετρική μορφή το σύστημα των σχέσεων I,J που υφίστανται μεταξύ των στοιχείων των δύο αυτών συνόλων, χρησιμοποιώντας τις ιδιότητες του Ευκλείδειου διανυσματικού χώρου R^N .

Η **παραγοντική ανάλυση**, πάνω στην οποία βασίζονται οι μέθοδοι της Ανάλυσης Δεδομένων, **αναπαριστά γραφικά διαφαινόμενες ή μη εκ των προτέρων σχέσεις** που υφίστανται μεταξύ των συνόλων I και J τις οποίες προβάλλει πάνω σ' ένα επίπεδο, με στόχο την ευκολότερη ερμηνεία των σχέσεων αυτών.

Με βάση αυτή τη διαδικασία, το κύριο χαρακτηριστικό των μεθόδων της ανάλυσης δεδομένων μπορεί να αποδοθεί με το εξής σλόγκαν:

«Αποδέχομαι μία απώλεια πληροφορίας προκειμένου να επιτύχω ένα όφελος σε σημασία»

Η προσέγγιση αυτή είναι καθαρά **ποιοτικού περιεχομένου** και επιτυγχάνεται με **αμιγείς αναλυτικές μεθόδους** και αποτελεί τη **νέα αντίληψη** που εισάγεται από τις μεθόδους της Ανάλυσης Δεδομένων.

ΠΑΡΟΥΣΙΑΣΗ
ΤΗΣ
ΠΑΡΑΓΟΝΤΙΚΗΣ ΑΝΑΛΥΣΗΣ
ΤΩΝ ΑΝΤΙΣΤΟΙΧΙΩΝ

Δημιουργία του Γάλλου καθηγητή
Jean Paul Benzecri (1932-2019)

(-Π.Α.Α-)

Οι πλέον βασικές μέθοδοι της Γαλλικής Σχολής είναι :

α) Η Παραγοντική Ανάλυση των Αντιστοιχιών

και

β) Η Ανιούσα Ιεραρχική Ταξινόμηση

Δημιουργός των οποίων είναι ο Γάλλος Καθηγητής J.P. Benzecri

Υπάρχουν βέβαια και άλλες μέθοδοι που συνθέτουν την Γαλλική Σχολή της Ανάλυσης Δεδομένων, όπως λ.χ:

-Η Ανάλυση σε Κύριες Συνιστώσες

-Η Ανάλυση των Τάξεων

-Η Διακριτική Παραγοντική Ανάλυση

-Η Συνεπαιγωγική Στατιστική

-Η Σημειομετρία

Η Παραγοντική Ανάλυση των Αντιστοιχιών

Η μαθηματική διατύπωση της συγκεκριμένης μεθόδου ανάλυσης οφείλεται στον Γάλλο καθηγητή ο οποίος δίδασκε στο Πανεπιστήμιο Pierre et Marie Curie PARIS VI **JEAN PAUL BENZECRI** (28 Φεβρουαρίου 1932 – 24 Νοεμβρίου 2019)

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.

Η γενική μορφή ενός πίνακα συμπτώσεων είναι:

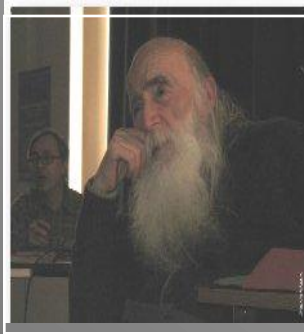
Πίνακας 2.1

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

Τα στοιχεία i του συνόλου I ονομάζονται «**άτομα**» ή «**αντικείμενα**», ενώ τα στοιχεία j του συνόλου J ονομάζονται «**μεταβλητές**». Στη διασταύρωση της γραμμής i και της στήλης j βρίσκεται ένας αριθμός που δηλώνει την σύμπτωση του στοιχείου i με την μεταβλητή j (απόλυτη συχνότητα) την οποία συμβολίζουμε με $k(i,j)$, ο δε πίνακας συμπτώσεων συμβολίζεται με K_{IJ} ή πίνακας $I \times J$.

Η ρήση του J.P BENZECRI

Ο J.P Benzecri, πρωτοπόρος της Γαλλικής Σχολής της Ανάλυσης Δεδομένων και εμπνευστής της Παραγοντικής Ανάλυσης των Αντιστοιχιών, στην αρχή του μαθήματος στο Πανεπιστήμιο Pierre et Marie Curie (Jussieu Paris VI) σχετικά με την παραγοντική ανάλυση ανέφερε τα εξής:

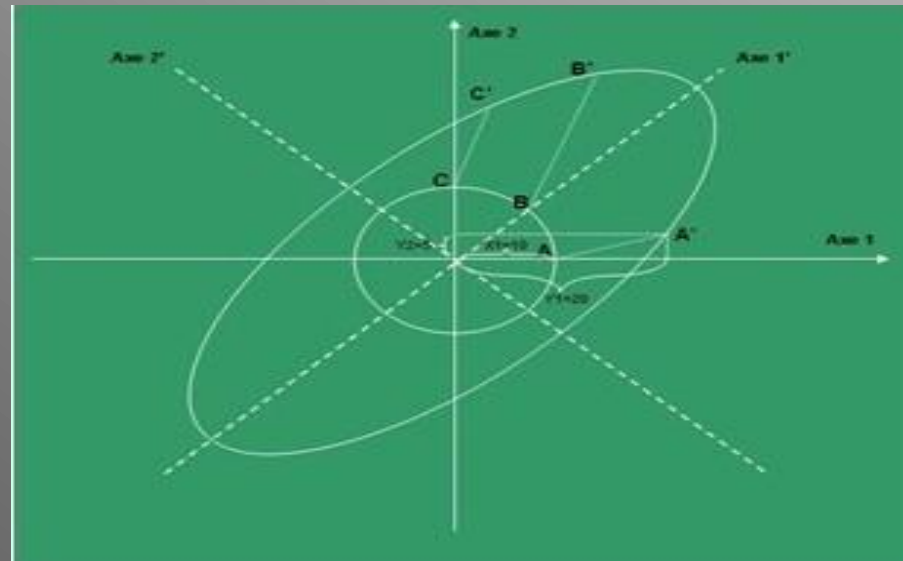


«Στη καρδιά κάθε ανάλυσης δεδομένων υπάρχει διαγωνοποίηση ενός συμμετρικού τετραγωνικού πίνακα».

Jean Paul Benzecri

Για τον Benzecri η Α.Δ συνιστά Φιλοσοφία, η οποία απελευθερώνει τον ερευνητή από δεσμεύσεις, που ενδεχομένως επιβάλλουν εξωγενείς παράγοντες, αφήνοντας τη φροντίδα και την ευθύνη να εξάγει ο ίδιος τις ερμηνείες των φαινομένων και τις συνέπειες που προκαλούν.

Αποτέλεσμα →



Οι άξονες 1 και 2 είναι οι άξονες του κύκλου ενώ οι άξονες 1' και 2' είναι άξονες της έλλειψης

Ο πίνακας των «προφίλ» των γραμμών f_j^i

Η βασική έννοια που δεσπόζει στην Παραγοντική Ανάλυση των Αντιστοιχιών είναι η έννοια του **προφίλ** είτε μιας γραμμής είτε μιας στήλης

Αρχικά ένας πίνακας δεδομένων μετατρέπεται σε πίνακα **σχετικών συχνοτήτων**. Για να είναι συγκρίσιμες δύο γραμμές i και i' του πίνακα 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 την παρακάτω σχέση

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

Όπου

k_{ij} = η απόλυτη συχνότητα της γραμμής i για την στήλη j

k_i = το άθροισμα των απολύτων συχνοτήτων της γραμμής i

ΠΑΡΑΤΗΡΗΣΗ: Όταν δύο γραμμές i και i' παρουσιάζουν **ανάλογες απόλυτες συχνότητες**, παρουσιάζουν το **ίδιο προφίλ**. Αυτό έχει ως συνέπεια συνήθως να συμπίπτουν πάνω στο παραγοντικό επίπεδο. Ως εκ τούτου όταν δύο προφίλ (γραμμές ή στήλες) εμφανίζονται στο παραγοντικό επίπεδο το ένα κοντά στο άλλο, σημαίνει ότι έχουν περίπου την ίδια σχετική κατανομή, άρα συμπεριφέρονται περίπου κατά τον ίδιο τρόπο, εκτός και αν βρίσκονται σε διαφορετικούς υποχώρους.

Αντίστοιχα ορίζονται και τα προφίλ των στηλών

Όπου k_j = το άθροισμα των απολύτων συχνοτήτων της στήλης j

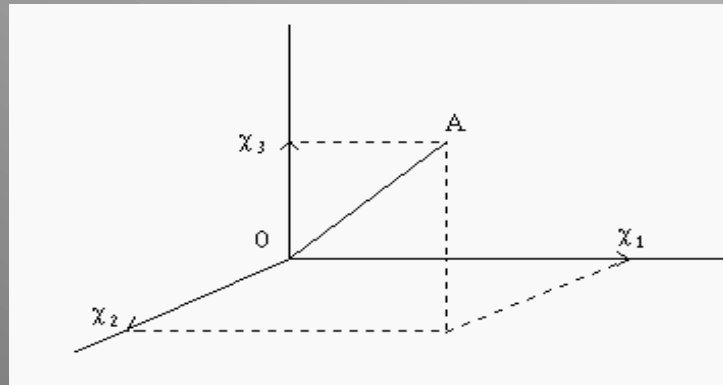
$$f_i^j = \frac{k_{ij}}{k_j}$$

Ευκλείδεια απόσταση

Στην Αναλυτική Γεωμετρία ένα ζεύγος πραγματικών αριθμών (X_1, X_2) ορίζει στο **Καρτεσιανό επίπεδο** ένα σημείο, ενώ μία τριάδα πραγματικών αριθμών (X_1, X_2, X_3) προσδιορίζει ένα σημείο στον **τριδιάστατο χώρο** *.

Επεκτείνοντας την αντίληψη αυτή μπορούμε να καθορίσουμε ότι μία σειρά n πραγματικών αριθμών, αποτελεί τις συντεταγμένες ενός σημείου στο **n -διάστατο χώρο**.

Ένα σημείο A στο τριδιάστατο χώρο παρουσιάζει το παρακάτω σχήμα



Στον n -διάστατο χώρο ισχύει το γενικευμένο **Πυθαγόρειο θεώρημα**

$$d^2(M_1, M_2) = \sum_{i=1}^n (x_2^i - x_1^i)^2$$

* Πολλοί χρησιμοποιούν τον όρο **τριδιάστατο**, λανθασμένα βέβαια, αφού ο συγκεκριμένος όρος **ΤΡΙΣ** προσδιορίζει, ότι υφίστανται **τρεις φορές η ίδια διάσταση** και **όχι τρεις διαφορετικές διαστάσεις** που αποδίδεται με τον ορθό για την συγκεκριμένη περίπτωση όρο **ΤΡΙ**. Κατά τη λανθασμένη αντίληψη, θα έπρεπε να προφέραμε τρίσκυκλο ποδήλατο αντί τρίκυκλο, τρίστομο έργο αντί τρίτομο κ.λ.π. Η ίδια παρατήρηση ισχύει και για τον λανθασμένο όρο **δισ**, αντί του ορθού **δι**. Π.χ με την ορθή προφορά αναφερόμαστε σε **διδιάστατο** χώρο και όχι με τον προφερόμενο λανθασμένο όρο **δισδιάστατο** χώρο. Βέβαια δεν πρέπει να παρασύρεται κάποιος από τους όρους τρισέλιδο, τρίστηλο, επειδή το γράμμα σ είναι το αρχικό των απλών λέξεων σελίδα, στήλη.

ΘΕΣΕΙΣ ΤΩΝ ΣΗΜΕΙΩΝ ΣΤΟ ΠΑΡΑΓΟΝΤΙΚΟ ΕΠΙΠΕΔΟ

Οι δυνατές "θέσεις" που μπορεί να παρουσιάσουν στο παραγοντικό επίπεδο τα διανύσματα τα οποία αντιπροσωπεύουν είτε στατιστικές μονάδες (γραμμές) είτε μεταβλητές (στήλες) είναι τρεις.

I) η θέση της συζυγίας ($0 < \theta_1 < 90$)

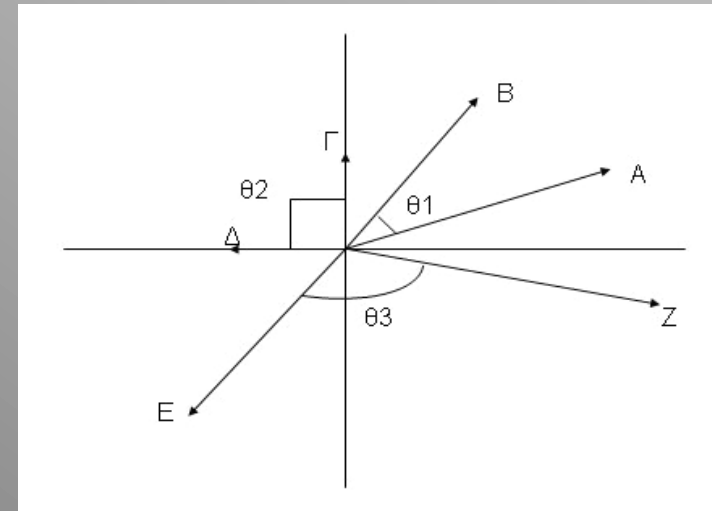
Σε κάθε συζυγία αντιστοιχεί θετική απόκλιση που εξομοιώνεται με **έλξη** των δύο διανυσμάτων. Η ένταση της συζυγίας εξαρτάται από το μήκος της προβολής και την γωνία θ_1 που σχηματίζουν τα δυο διανύσματα.

II) η θέση της καθετότητας ($\theta_2 = 90$)

Το περιεχόμενο του αντίστοιχου φατνίου είναι μηδέν, άρα δεν παρουσιάζει ούτε θετική ούτε αρνητική απόκλιση, ένδειξη που μας οδηγεί να θεωρήσουμε ότι τα δυο διανύσματα είναι **ανεξάρτητα** μεταξύ τους.

III) η θέση της αντίθεσης ($90 < \theta_3 < 180$)

Σε θέση αντίθεσης αντιστοιχεί η αρνητική απόκλιση που ερμηνεύεται ως **άπωση** μεταξύ των δύο διανυσμάτων.



Το νέφος $N(I)$ των σημείων που αντιπροσωπεύουν οι γραμμές ενός πίνακα

Σε κάθε σημείο I_i αντιστοιχούμε την **μάζα του m_i** , η οποία είναι ίση με την **περιθωριακή συχνότητα**

Το βαρύκεντρο βρίσκεται με βάση τον παρακάτω τύπο :

$$G_j = \frac{\sum_i m_i \cdot k_{ij}}{\sum_i m_i} \quad (j=1, \dots, p)$$

Αδράνεια του νέφους των στατιστικών μονάδων $N(I)$

Η παρουσιαζόμενη **αδράνεια $I(i, G)$** κάθε σημείου i ($i=1, \dots, n$), υπολογίζεται με το άθροισμα των τετραγώνων των αποκλίσεων των συντεταγμένων του σημείου i , από τις αντίστοιχες συντεταγμένες του βαρύκεντρου G , σταθμίζοντας κάθε απόκλιση με το βάρος m_i

όπου

$$I(i, G) = \sum_i m_i \|i - G\|^2$$

$$\|i - G\|^2 = \sum_j (i_j - G_j)^2 \quad \text{για } j=1, \dots, p$$

Η **συνολική αδράνεια** του νέφους $N(I)$ ισούται με το άθροισμα όλων των επί μέρους αδρανειών

$$I(N(i), G) = I_{\text{ολ}} = \sum_i I(i, G)$$

Ενώ η **διασπορά (διακύμανση)** του νέφους $N(I)$ υπολογίζεται από τη σχέση

$$\text{Var}(N) = \frac{I(N, G)}{\sum_i m_i}$$



Οι έννοιες αυτές **παραπέμπουν** ότι η μέθοδος της Παραγοντικής Ανάλυσης των Αντιστοιχιών **συνδέεται** άμεσα με τους νόμους της Φυσικής και της Συστημικής Προσέγγισης. Η προσέγγιση αυτή αποτελεί **ακρογωνιαίο λίθο** μιας νέας επιστήμης την **Οικονομική Φυσική** ή άλλως **EconoPhysics**

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

Η απόσταση $d_2(f_j^I, f_j^{I'})$ μεταξύ δύο προφίλ στηλών ορίζεται από τη σχέση

$$d^2(f_j^I, f_j^{I'}) = \sum_{j=1}^p \frac{1}{f_i} (f_j^I - f_j^{I'})^2$$

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

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

όπου

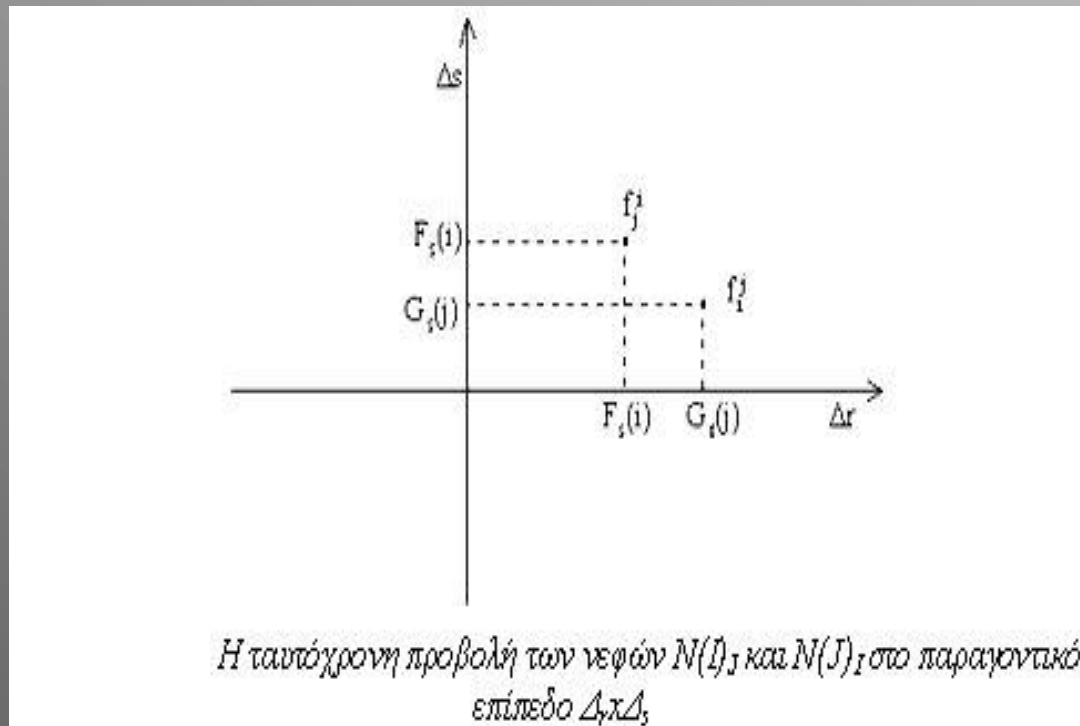
φ_a^j το χαρακτηριστικό διάνυσμα του παραγοντικού άξονα Δ_a

και

λ_a η χαρακτηριστική ρίζα που αντιστοιχεί στον παραγοντικό άξονα Δ_a

Τα παραγοντικά επίπεδα

Επειδή με την ανάλυση των νεφών $N(I)$ και $N(J)$ 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, προβάλλοντας τα 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.

ΠΑΡΟΥΣΙΑΣΗ
ΤΗΣ
ΑΝΙΟΥΣΑΣ ΙΕΡΑΡΧΙΚΗΣ
ΤΑΞΙΝΟΜΗΣΗΣ

Ανιούσα Ιεραρχική Ανάλυση

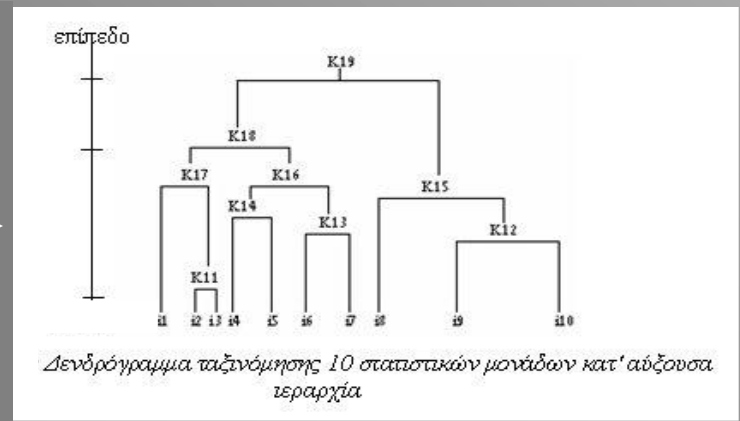
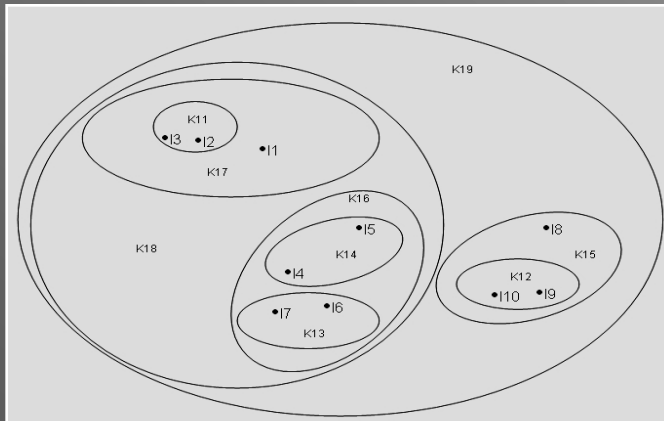
Μία Ανιούσα Ιεραρχική Ταξινόμηση (Classification Ascendante Hierarchique-CAH-) των στοιχείων ενός συνόλου I με πληθάρημο $\text{card}(I) = n$, είναι μία διαδικασία που παράγει μια **ακολουθία διαμελισμών** του αρχικού συνόλου σε **υποσύνολα μη κενά και ξένα ανά δύο μεταξύ τους**, τις λεγόμενες **κλάσεις**, **τη μία μέσα στην άλλη**, συνενώνοντας κάθε φορά **δύο μόνο κλάσεις** οι οποίες βάσει κάποιας μετρικής παρουσιάζουν σε κάθε βήμα ομαδοποίησης την **μικρότερη απόσταση**.

Όσο απομακρύνεται κανείς από τον αρχικό διαμελισμό (ο οποίος περιλαμβάνει τόσες κλάσεις όσα είναι τα αντικείμενα που ταξινομούνται), τόσο αυτός γίνεται λιγότερο λεπτομερής. Η τελευταία κλάση περιλαμβάνει το σύνολο των κλάσεων που δημιουργήθηκαν από τις συνενώσεις στοιχείων και κλάσεων, το πλήθος Π των οποίων είναι $2n-1$, όπου n το πλήθος των στατιστικών μονάδων που ταξινομούνται.

Η συνένωση δύο παρατηρήσεων σε μία κλάση δημιουργεί αυτό που ονομάζουμε **κόμβο** της ιεραρχίας.

Η αρίθμηση του 1ου κόμβου παίρνει τον αριθμό $n+1$ και συμβολίζεται με K_{n+1} . Π.χ Αρχικά έχουμε $n=10$ στοιχεία (κλάσεις) οπότε στο σύνολο δημιουργούνται $2 \times 10 - 1 = 19$ κλάσεις όπου οι κλάσεις K_1, \dots, K_{10} αντιστοιχούν στα 10 πρώτα στοιχεία που ταξινομούνται και έπεται η αρίθμηση των κόμβων.

Το σύνολο των εννέα κλάσεων $\{K_{11} \text{ έως } K_{19}\}$ ονομάζεται **τυπολογία** της ιεραρχίας



Δενδρόγραμμα ταξινόμησης 10 στατιστικών μονάδων κατ' αύξουσα ιεραρχία

Απόσταση μεταξύ των στατιστικών μονάδων

Ο πίνακας δεδομένων που υποβάλλεται σε CAH, είναι ένας πίνακας αποστάσεων $T(n \times n)$ μεταξύ των n στατιστικών μονάδων. Για τον υπολογισμό των αποστάσεων $d(i, i')$ δύο στατιστικών μονάδων υπάρχουν διάφορα είδη μετρικής, που χρησιμοποιούνται ανάλογα με την φύση των δεδομένων.

A) Για ποσοτικά δεδομένα

➤ Η Ευκλείδεια μετρική

$$d(i, i') = \sqrt{\sum_{j=1}^p [X(i, j) - X(i', j)]^2}$$

➤ Η μετρική του City-block (Manhattan)

$$d(i, i') = \sum_{k=1}^n |x_{ik} - x_{i'k}|$$

➤ Η μετρική του Mahalanobis

$$d(i, i') = \sqrt{(x_i - x_{i'}) \cdot C^{-1} \cdot (x_i - x_{i'})}$$

Όπου C η μήτρα Διακυμάνσεων - Συνδιακυμάνσεων

B) Για ποιοτικά δεδομένα

➤ Η μετρικής του χ^2

$$d^2(i, i') = \sum_{j=1}^p \frac{1}{P_j} (P_j^i - P_j^{i'})^2$$

Γ) όταν τα δεδομένα είναι ένας πίνακας συσχετίσεων

$$d(i, i') = \sqrt{1 - r_{ij}}$$

όπου το r_{ij} είναι ο συντελεστής συσχέτισης μεταξύ των μεταβλητών X_i και X_j .

Κριτήρια συνένωσης δύο κλάσεων

α) το κριτήριο της ελάχιστης απόστασης ή κριτήριο MIN

$$d(i \cup i', j) = \min\{d(i, j), d(i', j)\}$$

β) το κριτήριο της μέγιστης απόστασης ή κριτήριο MAX

$$d(i \cup i', j) = \max\{d(i, j), d(i', j)\}$$

γ) το κριτήριο MINMAX

$$d(i \cup i', j) = \min\{\max[d(i, j), d(i', j)]\}$$

δ) το κριτήριο της μέσης απόστασης

$$d(i \cup i', j) = \frac{p(i)d(i, j) + p(i')d(i', j)}{p(i) + p(i') + p(j)}$$

ε) το κριτήριο της απώλειας της διαταξικής αδράνειας

Κατά την διαδικασία της ταξινόμησης η αδράνεια του νέφους των σημείων των παρατηρήσεων χωρίζεται με βάση το θεώρημα του Huygens σε δύο συνιστώσες, την εσωταξική αδράνεια I_E και την διαταξική αδράνεια I_Δ .

$$I_{\text{ολ}} = I_E + I_\Delta$$

Διαδικασία δημιουργίας της ιεραρχίας. Κριτήριο του Ward

Το κριτήριο ομαδοποίησης είναι να συγχωνεύουμε δύο κλάσεις **A** και **B**, όταν η παρουσιαζόμενη απώλεια της διαταξικής αδράνειας μεταξύ δυο κλάσεων είναι η ελάχιστη.

Ο αλγόριθμος της μεθόδου

Η διαδικασία δημιουργίας ενός δενδρογράμματος ανιούσας ιεραρχικής ταξινόμησης ενός συνόλου N «αντικειμένων» παρουσιάζει έξι βήματα.

1ο βήμα: Κάθε στοιχείο ενός συνόλου N «αντικειμένων» προς κατάταξη θεωρείται ότι αποτελεί μία κλάση, συνεπώς η αρίθμηση των κόμβων πριν την δημιουργία της πρώτης κλάσης, αρχίζει από το 1 και καταλήγει στο N . Στο βήμα αυτό υπολογίζονται τα προφίλ των «αντικειμένων».

2ο βήμα: Υπολογίζονται οι αποστάσεις όλων των προφίλ των «αντικειμένων» μεταξύ τους.

3ο βήμα: Μετατρέπονται οι αποστάσεις του 2ου βήματος σε απώλειες διαταξικής αδράνειας.

4ο βήμα: Συνενώνονται τα «αντικείμενα» που παρουσιάζουν την ελάχιστη απώλεια αδράνειας. Ο κόμβος που δημιουργείται παίρνει την αρίθμηση K_{N+1} . Υπολογίζεται ο δείκτης « δ » του κόμβου.

5ο βήμα: Υπολογίζονται οι απώλειες αδράνειας της νέας κλάσης K_{N+1} με τα υπόλοιπα $N-1$ «αντικείμενα».

6ο βήμα: Επαναλαμβάνονται τα βήματα 4 και 5 μέχρι να καταλήξουμε σε μία κλάση που περιλαμβάνει και τα N «αντικείμενα». Η αρίθμηση του κόμβου που αντιστοιχεί στην τελευταία κλάση της ανιούσας ιεραρχίας είναι $2 \cdot N - 1$.

Συμβολή του Καθηγητή Δρος Δημητρίου Ν. Καραπιστόλη στην εξέλιξη της μεθόδου του δασκάλου του Jean Paul Benzecri

ΕΠΙΣΗΜΑΝΣΕΙΣ του καθηγητή Δ. Καραπιστόλη

Τα βασικά χαρακτηριστικά που παρουσιάζει η Παραγοντική Ανάλυση των Αντιστοιχιών τα οποία κατά την επιστημονική άποψη του καθηγητή Δ. Καραπιστόλη χρήζουν βελτίωσης είναι τα εξής:

a) Κατ' αρχή η **Παραγοντική Ανάλυση των Αντιστοιχιών** λαμβάνει υπόψη, όπως αναφέρει ο J.P. Benzecri, τις ερμηνείες έως τους πρώτους επτά παραγοντικούς άξονες, από το πλήθος των $n-1$ αξόνων, διατυπώνοντας το σλόγκαν

«Αποδέχομαι μία απώλεια πληροφορίας προκειμένου να επιτύχω ένα όφελος σε σημασία»

b) Για την καλύτερη ερμηνεία των σχέσεων που απεικονίζονται κυρίως στο 1^ο παραγοντικό επίπεδο 1×2 , ο οποίος συγκεντρώνει την μεγαλύτερη αδράνεια, χρειάζεται να πραγματοποιηθεί στα δεδομένα **χωριστά η Ανιούσα Ιεραρχική Ταξινόμηση**, προς αποφυγή παρερμηνειών δύο γειτονικών σημείων των οποίων οι προβολές βρίσκονται στο ίδιο επίπεδο, αλλά μπορεί να ανήκουν σε διαφορετικούς υποχώρους.

c) Σχετικά με τα αποτελέσματα της **Ανιούσας Ιεραρχικής Ταξινόμησης** μία κλάση K_i αφορά αποκλειστικά ένα υποσύνολο «ατόμων» με **παρόμοια συμπεριφορά** ως προς το σύνολο των μεταβλητών της ανάλυσης, όπου με βάση την μέση τιμή κάθε μεταβλητής προκύπτει η συμβολή της στη διαμόρφωσή της κλάσης. Αυτό έχει ως αποτέλεσμα να μην γνωρίζει ο ερευνητής εκ των προτέρων την **βαρύτητα** κάθε μεταβλητής, η οποία λόγω του διαφορετικού πλήθους «ατόμων» σε κάθε κλάση είναι διαφορετική για τις κλάσεις της τυπολογίας.

d) Ως απόρροια της 3^{ης} παρατήρησης, μετά την δημιουργία της τυπολογίας της ταξινόμησης δεν μπορεί να γίνει **ΚΑΜΙΑ ΑΞΙΟΛΟΓΗΣΗ της ΟΜΟΙΟΓΕΝΕΙΑΣ των κλάσεων**, κάτι που εκ των πραγμάτων έπρεπε να συμβαίνει. Στο σημείο αυτό πρέπει να τονιστεί πως ούτε η μηχανή μάθησης **SVM**, η καλύτερη από όσες κυκλοφορούν, δεν έχει την δυνατότητα **να υπολογίσει την πιθανότητα συμμετοχής** ενός ατόμου στην κλάση που κατατάχθηκε, στοιχείο που θα προσδιόριζε την ζητούμενη ομοιογένεια.

Για να αντιμετωπιστούν οι αναφερόμενες προφανείς διαπιστώσεις, ο Δρ. Δημήτριος Νικ. Καραπιστόλης επέκτεινε τα αποτελέσματα της Παραγοντικής Ανάλυσης των Αντιστοιχιών και της Ανιούσας Ιεραρχικής Ταξινόμησης, δημιουργώντας νέες μεθόδους, όπως την **μέθοδο KARAP**, την **μέθοδο BENKAR**, την **Σύνθετη Ιεραρχική Ταξινόμηση**, την **μέθοδο της Αξιολόγησης των κλάσεων** της Ανιούσας Ιεραρχικής Ταξινόμησης

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΟΛΟ ΣΗΜΕΙΩΜΑ

Ο καθηγητής Δρ. Δημήτριος Νικ. Καραπιστόλης υπηρέτησε για 38 χρόνια την Ανώτατη Εκπαίδευση.

Κατά την περίοδο 13/12/1976-14/5/1980 υπηρέτησε στο νυν Πανεπιστήμιο Μακεδονίας ως βοηθός στη έδρα της Στατιστικής

Στη συνέχεια την περίοδο από 15-5-1980 έως 31-8-14 διετέλεσε καθηγητής Στατιστικής και Ανάλυσης Δεδομένων στο τμήμα Εμπορίας και Διαφήμισης του Α.Τ.Ε.Ι Θεσσαλονίκης, νυν Διεθνές Πανεπιστήμιο Ελλάδος (ΔΙΠΑΕ)

A. ΣΠΟΥΔΕΣ-ΤΙΤΛΟΙ

A.1 Διδάκτωρ του Τμήματος Εφαρμοσμένης Πληροφορικής του Πανεπιστημίου Μακεδονίας Οικονομικών και Κοινωνικών Επιστημών. Ο τίτλος της διατριβής του είναι «**Δημιουργία λογισμικού για την κατάρτιση φερέγγυου χαρτοφυλακίου με μεθόδους ανάλυσης δεδομένων**». Το γνωστικό αντικείμενο της διατριβής εντάσσεται στα επιστημονικά πεδία της Στατιστικής, της Ανάλυσης Δεδομένων, της Διαχείρισης χαρτοφυλακίων μετοχών εισηγμένων εταιρειών στο χρηματιστήριο. Βαθμός ΑΡΙΣΤΑ

A.2 Διπλωματούχος του Τμήματος Μεταπτυχιακών σπουδών του κλάδου Οικονομικών Επιστημών της Σχολής Νομικών και Οικονομικών Επιστημών του Αριστοτέλειου Πανεπιστημίου Θεσ/νίκης. Βαθμός ΑΡΙΣΤΑ

A.3 Πτυχιούχος Μαθηματικός της Φυσικομαθηματικής Σχολής του Αριστοτέλειου Πανεπιστημίου Θεσσαλονίκης. Ημερομηνία λήψης του πτυχίου 24 ΦΕΒΡΟΥΑΡΙΟΥ 1973. Βαθμός ΛΙΑΝ ΚΑΛΩΣ.

B. ΕΠΙΣΤΗΜΟΝΙΚΟ ΕΡΓΟ

Το επιστημονικό έργο του καθηγητή περιλαμβάνει την διδακτορική διατριβή, τρία λογισμικά (το **MAD** που αφορά αναλύσεις δεδομένων με μεθόδους πολυπαραγοντικής στατιστικής ανάλυσης, το **PPM** που αφορά μελέτη χρηματιστηριακών μετοχών, το **MFM** που αφορά μελέτη Αμοιβαίων Κεφαλαίων βάσει δικής του πρωτότυπης μεθοδολογίας, **δέκα βιβλία και δύο μονογραφίες, δημοσιεύσεις** μετά από κρίση σε επιστημονικά περιοδικά, συμμετοχές **με ανακοινώσεις** σε διεθνή συνέδρια, **επιστημονικός υπεύθυνος** σε ερευνητικά προγράμματα και πλήθος αναφορών στη πλατφόρμα ResearchGate στο επιστημονικό του έργο.

Γ. Διοικητική δραστηριότητα . Επί 3 διετίες Προϊστάμενος του Τμήματος

Πηγή: www.pylimad.gr

ΠΡΟΤΕΙΝΟΜΕΝΕΣ ΒΕΛΤΙΩΣΕΙΣ

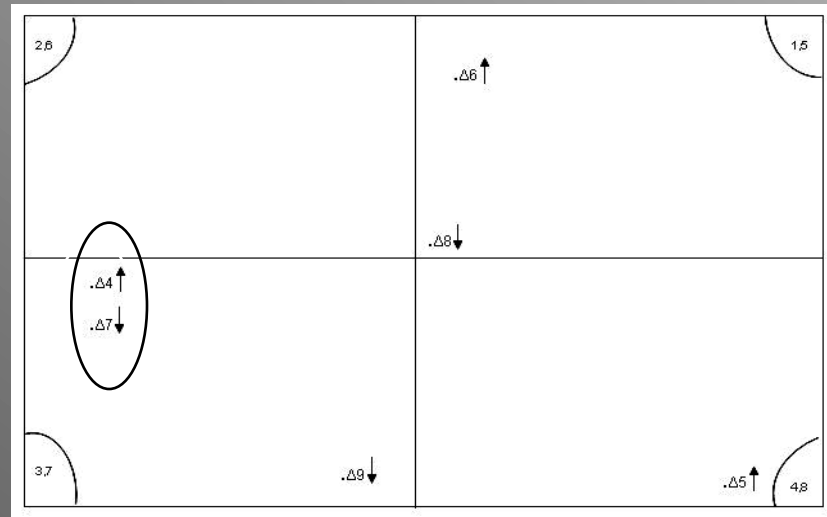
ΔΙΑΓΡΑΜΜΑΤΑ ΚΑΡΑΠΙΣΤΟΛΗ

Εφόσον χρησιμοποιηθούν οι τρεις πρώτοι παραγοντικοί άξονες και με δεδομένο ότι αποτελούν στον χώρο των τριών διαστάσεων ένα τρισσορθογώνιο σύστημα συντεταγμένων, τότε δημιουργούνται οκτώ υποχώροι, στους οποίους ανάλογα με τα πρόσημα των συντεταγμένων των σημείων τοποθετούνται σ' έναν απ' αυτούς.

Συνεπώς αν τοποθετήσουμε τα σημεία που βρίσκονται στο 1^ο, 2^ο, 3^ο και 4^ο υποχώρο το σύμβολο ↑ δίπλα από την ταυτότητα του σημείου, ενώ για τα σημεία στον 5^ο, 6^ο, 7^ο και 8^ο υποχώρο θέσουμε το σύμβολο ↓, έχουμε για πρώτη φορά στη παγκόσμια βιβλιογραφία, μια απεικόνιση του τριδιάστατου χώρου στο επίπεδο, (χωρίς να χρησιμοποιείται η μέθοδος της προοπτικής), οπότε τα διαγράμματα αυτά στο εξής θα αναφέρονται ως **διαγράμματα Καραπιστόλη**.

Συνεπώς με τα διαγράμματα Καραπιστόλη απεικονίζεται **παραστατικά ο τριδιάστατος υποχώρος στο επίπεδο**, ώστε ο ερευνητής να μην χρειάζεται να πραγματοποιήσει ταξινόμηση (εφόσον βέβαια η ερμηνεία των τριών πρώτων παραγοντικών αξόνων είναι ικανοποιητική δηλαδή πάνω του 75%), κατά συνέπεια να μην συγχέει γειτονικά σημεία τα οποία ανήκουν σε διαφορετικούς υποχώρους του R^3

Παραστατικά έχουμε



Από το **διάγραμμα Καραπιστόλη** διαπιστώνεται ότι οι μεταβλητές Δ4 (στο 3^ο και Δ7 στο 7^ο υποχώρο), καθώς και οι μεταβλητές Δ6 και Δ8 στον τριδιάστατο χώρο βρίσκονται σε διαφορετικούς υποχώρους του, με ότι αυτό μπορεί να σημαίνει για την ερμηνεία τους σε ποσοστό 80,11%, έναντι της ερμηνείας τους σε ποσοστό 63,93% που παρέχει το παραγοντικό επίπεδο 1x2 (όπως εμφανίζονται τα σημεία χωρίς τα βέλη), δηλαδή μία απώλεια πληροφορίας σε ποσοστό 17,18% .

ΜΕΘΟΔΟΣ KARAP

A) KARAP vs FACOR-VACOR

Σχετικά με τα αποτελέσματα της **Ανιούσας Ιεραρχικής Ταξινόμησης** μία κλάση K_i όπως προαναφέρθηκε, αφορά αποκλειστικά ένα υποσύνολο «ατόμων» με παρόμοια συμπεριφορά ως προς το σύνολο των μεταβλητών, καθόσον οι αποστάσεις των σημείων («ατόμων») υπολογίζονται με βάση τα προφίλ τους, κάνοντας χρήση της **μετρικής χ^2** (**απόσταση Benzecri**)

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

Αυτό έχει ως αποτέλεσμα να μην γνωρίζει ο ερευνητής το κάθε «άτομο» της κλάσης K_i με ποια μεταβλητή **ΣΥΝΔΕΕΤΑΙ** περισσότερο, με άλλα λόγια το προφίλ του με ποιο προφίλ μιας μεταβλητής είναι **ΕΓΓΥΤΕΡΟ**. Αν τώρα όλα τα άτομα που συμμετέχουν στην κλάση K_i είναι συνδεδεμένα με συγκεκριμένη μεταβλητή, τότε μπορούμε να υποστηρίξουμε, ότι η μεταβλητή αυτή χαρακτηρίζει την συγκεκριμένη κλάση K_i

Η μέθοδος KARAP απαντά στον συγκεκριμένο προβληματισμό που υφίσταται σε κάθε Ανιούσα Ιεραρχική Ταξινόμηση είτε αυτή δημιουργείται με τη μέθοδο FACOR, είτε με τη μέθοδο VACOR, ότι δηλαδή εντοπίζονται με ακρίβεια τα αντικείμενα που συνδέονται με την κάθε μεταβλητή. Για την επίτευξη αυτού του στόχου η μέθοδος KARAP χρησιμοποιεί την **Ευκλείδειο απόσταση στον ν-διάστατο Ευκλείδειο διανυσματικό χώρο**.

$$d(i, i') = \sqrt{\sum_{j=1}^p [X(i, j) - X(i', j)]^2}$$

Η διαδικασία που ακολουθείται προβλέπει να εντοπιστούν όλα τα «άτομα» που συνδέονται με κάθε μεταβλητή χωριστά που παρουσιάζουν την μικρότερη απόσταση με αυτή. Συνεπώς επιτυγχάνει στο **μέγιστο βαθμό** που επιθυμεί ο ερευνητής, την **ομοιογένεια** των αντικειμένων ως προς την αποκλειστική σύνδεσή τους με κάθε μία μεταβλητή.

Με την χρήση, λοιπόν, της μεθόδου KARAP επιτυγχάνουμε την **αμιγή σύνθεση** των κλάσεων της Ανιούσας Ιεραρχίας, ως προς την συμμετοχή αντικειμένων με καθορισμένα χαρακτηριστικά των μεταβλητών που συμβάλλουν καθοριστικά στη δημιουργία των κλάσεων αυτών, πληροφορία η οποία **δεν επιτυγχάνεται** ούτε με την διαδικασία FACOR ούτε με τη διαδικασία VACOR.

Β) Ταξινόμηση KARAP με βάση τον πλησιέστερο συγγενή

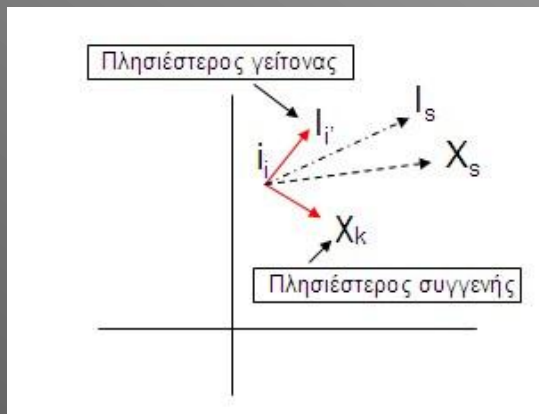
Η προτεινόμενη μέθοδος ταξινόμησης με το όνομα «**Μέθοδος Ταξινόμησης KARAP**» (**Nearest Kinsman Classification-NKC-**) διαφέρει από τη διαδικασία ταξινόμησης με την μέθοδο VACOR η οποία βασίζεται στον όρο «**πλησιέστερος γείτονας**» του J.P. Benzecri, όπου κατά τη διαδικασία της συνένωσης δύο κλάσεων «αντικειμένων» χρησιμοποιεί την μέθοδο Ward, με κριτήριο την ελάχιστη απώλεια της διαταξικής αδράνειας και μετρική την «απόσταση X^2 » που θεσπίστηκε από τον J.P. Benzecri.

Η διαφορά έγκειται ότι στη διαδικασία NKC η οποία προτείνεται από τον Δρ. Δημήτριο Καραπιστόλη, χρησιμοποιεί τον όρο «**πλησιέστερος συγγενής**», για την δημιουργία της ταξινόμησης των «αντικειμένων» η οποία είναι συνεκτική ως προς την μεταβλητή με την οποία συνδέονται συγκεκριμένα «αντικείμενα», ενώ στη διαδικασία συνένωσης δύο κλάσεων των μεταβλητών χρησιμοποιείται η μέθοδος WARD με κριτήριο την ελάχιστη διακύμανση εντός των κλάσεων, με βάση τον αλγόριθμο των Lance-Williams χρησιμοποιώντας την Ευκλείδεια μετρική στον πίνακα των συντεταγμένων των μεταβλητών πάνω στους $p-1$ άξονες που προκύπτουν από την εφαρμογή της Παραγοντικής Ανάλυσης των Αντιστοιχιών στα στοιχεία του πίνακα δεδομένων προς ταξινόμηση.

Ειδικότερα:

Ένα «αντικείμενο» I_i θεωρεί ως «**πλησιέστερο γείτονα**» τον I_j , όταν το προφίλ του I_i έχει με το προφίλ του I_j την μικρότερη απόσταση από κάθε άλλο «αντικείμενο» I_s του δείγματος, με βάση την **απόσταση x^2** που θεσπίστηκε από τον J.P. Benzecri.

Ο νέος όρος που θεσπίζεται θεωρεί για ένα «αντικείμενο» I_i ως «**πλησιέστερο συγγενή**» την μεταβλητή X_k όταν το προφίλ του I_i έχει με το προφίλ της μεταβλητής X_k την μικρότερη απόσταση από κάθε άλλη μεταβλητή X_s του δείγματος με βάση την **μέθοδο KARAP**.



Τα βήματα της μεθόδου

Βήμα 1ο : Ο πίνακας δεδομένων $T(n,p)$ αναλύεται με την Παραγοντική Ανάλυση των Αντιστοιχιών

Βήμα 2ο : Με τη χρήση της μεθόδου **KARAP**, για κάθε «αντικείμενο» εντοπίζεται ο «πλησιέστερος συγγενής» του με βάση την μετρική X^2

Βήμα 3ο : Γίνεται η εξαγωγή του αρχείου των $p-1$ συντεταγμένων των μεταβλητών GA, σε μορφή αρχείου .TXT. Θα χρησιμοποιηθεί για την κατασκευή της ταξινόμησης των μεταβλητών με βάση την Ευκλείδειο μετρική (λόγω του ορθοκανονικού συστήματος που δημιουργούν) και την μέθοδο της ελάχιστης διακύμανσης εντός των κλάσεων του WARD (Αγγλοσαξονική διαδικασία cluster), αντί της διαδικασίας της ελάχιστης απώλειας της διαταξικής αδράνειας (Γαλλική διαδικασία με χρήση της μετρικής X^2)

Βήμα 4ο : Χρησιμοποιώντας την κατάταξη των «αντικειμένων» σε συνδυασμό με την ταξινόμηση των μεταβλητών, προκύπτει ένα πλήρες δενδρόγραμμα, όπου έχουμε την ταυτόχρονη παρουσία της σύνδεσης των «αντικειμένων» με τις αντιπροσωπευτικές μεταβλητές με τις οποίες συσχετίζονται περισσότερο (βλέπε σχήμα 1).

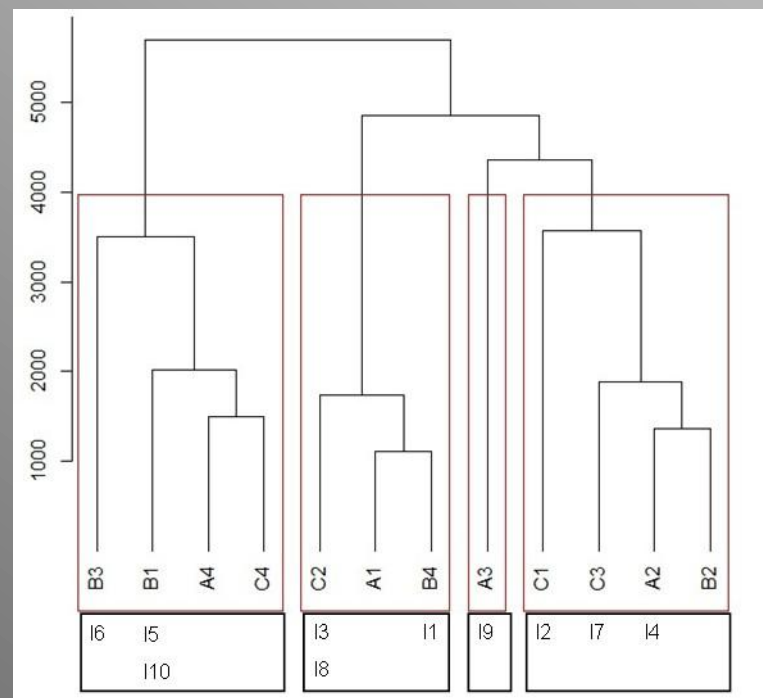
Ο λογικός πίνακας 0-1 των 3 μεταβλητών

INT	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
I1	1	0	0	0	0	0	0	1	0	0	1	0
I2	0	1	0	0	0	1	0	0	1	0	0	0
I3	1	0	0	0	0	0	0	1	0	1	0	0
I4	0	1	0	0	0	1	0	0	0	0	1	0
I5	1	0	0	0	1	0	0	0	0	0	0	1
I6	0	0	0	1	0	0	1	0	0	0	0	1
I7	0	1	0	0	0	0	0	1	0	0	1	0
I8	1	0	0	0	0	0	0	1	0	1	0	0
I9	0	0	1	0	0	1	0	0	0	0	0	1
I10	0	0	0	1	1	0	0	0	0	0	0	1

Οι θέσεις των μεταβλητών και των «αντικειμένων» με βάση τα αποτελέσματα της μεθόδου KARAP

A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
	I4	I9		I5		I6	I1	I2	I3	I7	
				I10					I8		

Το δενδρόγραμμα της NCK



σχήμα 1

ΜΕΘΟΔΟΣ ΒΕΝΚΑΡ

1. Χρησιμοποιούμε τις συντεταγμένες του συνόλου των σημείων των νεφών $N(I)$ και $N(J)$, καθώς επίσης τις συντεταγμένες των κόμβων k (κλάσεις) της ταξινόμησης στους $p-1$ παραγοντικούς άξονες που προκύπτουν από την ανάλυση του πίνακα $T(n+k,p)$.
2. Ακολουθώντας με την χρήση της Ευκλείδειας μετρικής μπορούμε να υπολογίσουμε τις αποστάσεις της κάθε στατιστικής μονάδας από τα k κέντρα των κλάσεων
3. Οι k αποστάσεις της κάθε στατιστικής μονάδας μετατρέπονται σε k πιθανότητες, όπου η πιο μικρή από τις k αποστάσεις, αντιστοιχεί στη μεγαλύτερη πιθανότητα που έχει η στατιστική μονάδα να είναι πλησίον του κέντρου της κλάσης με την μικρότερη απόσταση.
4. Με τον τρόπο αυτό αντιμετωπίζεται το μειονέκτημα του αλγορίθμου Ward, όπου νομοτελειακά ακραίες τιμές κάποιων μεταβλητών θα συμμετέχουν **ΟΠΩΣΔΗΠΟΤΕ** σε κάποιες κλάσεις, αλλοιώνοντας την **ομοιογένειά** τους, που είναι το βασικό ζητούμενο σε μία ταξινόμηση.
5. Χρησιμοποιώντας τις μέγιστες πιθανότητες των «ατόμων» δημιουργούμε μία κατανομή πιθανοτήτων, συνήθως σε ΠΕΝΤΕ κλάσεις, απ' όπου προκύπτει η ζητούμενη αξιολόγηση της αρχικής ταξινόμησης με την μέθοδο FACOR, κάνοντας χρήση της κατανομής DIS (η οποία προέκυψε με βάση την Ευκλείδεια μετρική) ως **ενδιάμεση** κατανομή σύγκρισης της ορθής τοποθέτησης των n στατιστικών μονάδων εντός των 5 κλάσεων. Αν βέβαια το ποσοστό των αντικειμένων των δύο τελευταίων τάξεων είναι ικανοποιητικό, σε συνάρτηση με το ποσοστό που προσδιορίζει το εύρος των δύο τελευταίων τάξεων της κατανομής πιθανοτήτων, καθορίζει την αξιολόγηση της ταξινόμησης με την μέθοδο FACOR,
6. Η λογική της χρήσης της κατανομής DIS παρουσιάζεται στη περιγραφή της μεθόδου BENKAR και ιδιαίτερα στην παράγραφο «εκπαίδευση δεδομένων με τη Μηχανή Διανυσμάτων Υποστήριξης –ΜΔΥ-», όπου αποδεικνύεται ότι η ταξινόμηση των «ατόμων» με την Ευκλείδεια μετρική είναι περισσότερο κατανοητή, από εκείνη που πραγματοποιείται με την μέθοδο FACOR ή την μέθοδο VACOR με βάση τον αλγόριθμο του Ward.

11 0 ΠΑΝΕΛΛΗΝΙΟ ΣΥΝΕΔΡΙΟ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ

ΓΡΕΒΕΝΑ 29 Σεπτεμβρίου - 2 Οκτωβρίου 2022

ΣΥΝΘΕΤΗ ΙΕΡΑΡΧΙΚΗ ΤΑΞΙΝΟΜΗΣΗ

Δρ. Δημήτριος Καραπιστόλης

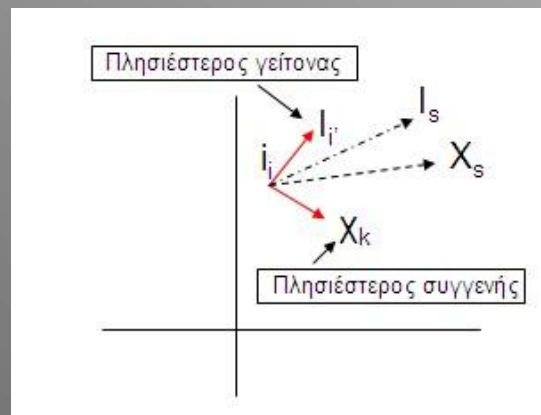
ΕΙΣΑΓΩΓΗ ΣΤΗ ΜΕΘΟΔΟ

Η προτεινόμενη μέθοδος ταξινόμησης με το όνομα «**Σύνθετη Ιεραρχική Ταξινόμηση**» (**Complex Hierarchical Classification-CHC-**) διαφέρει από τη διαδικασία ταξινόμησης με την μέθοδο VACOR, η οποία βασίζεται στον όρο «**πλησιέστερος γείτονας**» του J.P. Benzecri, ως προς το νέο όρο «**πλησιέστερος συγγενής**» που χρησιμοποιείται στη διαδικασία CHC από τον καθηγητή Δρ. Δημήτριο Καραπιστόλη.

Ο μεν πρώτος όρος θεωρεί για ένα «αντικείμενο» I_i ως «**πλησιέστερο γείτονα**» τον I_j , όταν το προφίλ του I_i έχει με το προφίλ του I_j την μικρότερη απόσταση από κάθε άλλο «αντικείμενο» I_s του δείγματος, με βάση την **απόσταση x^2** που θεσπίστηκε από τον J.P. Benzecri.

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

Ο νέος όρος θεωρεί για ένα «αντικείμενο» I_i ως «**πλησιέστερο συγγενή**» την μεταβλητή X_k όταν το προφίλ του I_i έχει με το προφίλ της μεταβλητής X_k την μικρότερη απόσταση από κάθε άλλη μεταβλητή X_s του δείγματος με βάση την **μέθοδο KARAP**.



Σκοπός της προτεινόμενης μεθόδου είναι αφενός να αξιολογήσει την **ενιαία σημαντικότητα των μεταβλητών για το σύνολο των κλάσεων** της διαμορφούμενης τυπολογίας της ιεραρχίας, αφετέρου να προκύψει μία φθίνουσα ιεραρχική ταξινόμηση των «αντικειμένων», η οποία στη συνέχεια μπορεί να χρησιμοποιηθεί ως διαμελισμός του συνόλου των «αντικειμένων», σε όσες κλάσεις κριθεί σκόπιμο από τον αναλυτή, ανάλογη διαδικασία με εκείνη που ισχύει με την τομή του δένδρογράμματος στη μέθοδο της Ανιούσας Ιεραρχικής Ταξινόμησης.

CHC vs VACOR

Ως γνωστόν στη μέθοδο VACOR οι μεταβλητές συμμετέχουν στη διαμόρφωση των κλάσεων μέσω των προφίλ των γραμμών, όπου η γειτνίαση δύο γραμμών (πλησιέστερων γειτόνων) βάσει της απόστασης $d^2(i_1, i_2)$ (απόσταση X^2) μετατρέπεται σε **απώλεια της διαταξικής αδράνειας** $\delta(i_1, i_2)$.

Ο υπολογισμός της διαταξικής αδράνειας I_Δ γίνεται με την χρήση του τύπου (1)

$$I_\Delta = \sum_i m_i \cdot d^2(g_i, G) \quad (1)$$

όπου με m_i συμβολίζουμε το συνολικό βάρους των στατιστικών μονάδων της K_i κλάσης, με g_i το βαρύκεντρο της κλάσης K_i και G το βαρύκεντρο του νέφους $N(I)$.

Η συνένωση δύο κλάσεων προκύπτει από την ελάχιστη απώλεια της διαταξικής αδράνειας, που προκαλείται με την συμμετοχή ενός νέου στοιχείου στην υπό διαμόρφωση κλάση.

Ο τύπος τώρα που δίνει την «**απόσταση**» $\delta(\Gamma, A \cup B)$ μεταξύ της κλάσης Γ και της κλάσης $A \cup B$ που προήλθε από την συγχώνευση των κλάσεων A και B έχει τη παρακάτω μορφή

$$\delta(\Gamma, A \cup B) = \frac{(p_A + p_\Gamma) \cdot \delta(A, \Gamma) + (p_B + p_\Gamma) \cdot \delta(B, \Gamma) - p_\Gamma \cdot \delta(A, B)}{p_A + p_B + p_\Gamma} \quad (2)$$

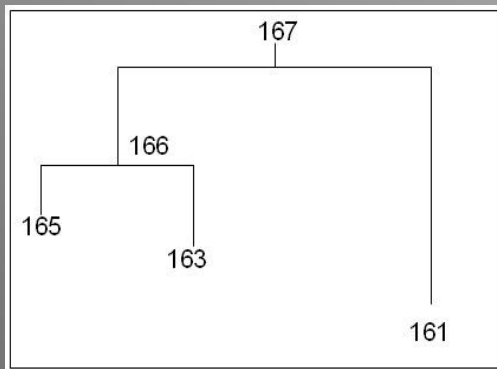
Η μέθοδος VACOR εκτός του ότι παρέχει το δενδρόγραμμα της ταξινόμησης μέσω της διαμορφούμενης τυπολογίας των κλάσεων, παρέχει επίσης τη δυνατότητα να προσδιοριστεί το ποσοστό συμβολής της κάθε μεταβλητής X_j , βάσει του οποίου χαρακτηρίζει κάθε κλάση K_i

Έτσι **ανάλογα με το πλήθος** των «αντικειμένων» που συμμετέχουν σε κάθε κλάση K_i υπολογίζεται η **μέση ποσοστιαία τιμή** κάθε μεταβλητής X_j ($j=1, \dots, p$), οπότε οι μεταβλητές που παρουσιάζουν **υψηλά** ή **χαμηλά** ποσοστά των τιμών τους, χαρακτηρίζουν την κλάση K_i με την έντονη **παρουσία** ή **απουσία** τους αντίστοιχα.

Συνεπώς η **σημαντικότητα** των μεταβλητών δεν είναι **ενιαία** για όλες τις κλάσεις, αλλά διαφορετική σε κάθε κλάση χωριστά.

Η σημαντικότητα των μεταβλητών στη μέθοδο VACOR

Θεωρούμε ένα πίνακα με 84 γραμμές και 6 μεταβλητές. Ο χώρος των κόμβων που δημιουργείται είναι ο R^6 , ενώ δημιουργούνται $2 \cdot 84 - 1 = 167$ κλάσεις. Η 1η διάσπαση της κλάσης 167 δημιουργεί το δίπολο των κλάσεων 166-161, ενώ η αμέσως επόμενη διάσπαση της κλάσης 166, δημιουργεί το δίπολο 165-163



Εικόνα 1: Δενδρόγραμμα της ταξινόμησης

Τυπολογία σε τρεις κλάσεις

Η πρώτη στήλη (#) κάθε κόμβου αντιπροσωπεύει τη συντεταγμένη του κέντρου της κλάσης στον άξονα που δημιουργεί η κάθε μεταβλητή.

Όσο αφορά στη στήλη COR(v) αντιπροσωπεύει γενικά τη **σχετική συμβολή** της μεταβλητής X_v στην απόκλιση της κλάσης K_i από το κέντρο του νέφους $K_{(2n-1)}$, οπότε για την μεταβλητή $\Delta 1$ έχουμε $COR_{166(\Delta 1)} = 75$, που σημαίνει ότι η μεταβλητή $\Delta 1$ συμβάλει στη **ΑΠΟΚΛΙΣΗ** της κλάσης 166 από τον κόμβο 167 σε ποσοστό 75%, το οποίο ερμηνεύεται και ως το ποσοστό που **χαρακτηρίζει** η μεταβλητή $\Delta 1$ την κλάση 166 στη διάσπαση της κλάσης 167 δημιουργώντας το δίπολο 166-161, άρα και την **σημαντικότητα** της μεταβλητής $\Delta 1$ για την κλάση 166.

Σημαντική παρατήρηση: Ο χαρακτηρισμός της $\Delta 1$ (όπως και των άλλων μεταβλητών) που αφορά την κλάση 166 είναι **διαφορετικός** για κάθε κλάση της ιεραρχίας. Συνεπώς **ΔΕΝ** αποτελεί ένα **ενιαίο χαρακτηρισμό για το σύνολο των κλάσεων της ταξινόμησης**, αλλά αφορά **μόνο κάθε κλάση χωριστά**, στοιχείο που αποτελεί την **βασική διαφορά** στη διαδικασία ταξινόμησης αντικειμένων μεταξύ της μεθόδου VACOR και της προτεινόμενης μεθόδου CHC.

d84.cla																		
Συμβολή των μεταβλητών στον χαρακτηρισμό των κόμβων																		
Κόμβος	167			166			165			164			163			162		
A(I)	166			165			164			159			156			157		
B(I)	161			163			162			148			155			160		
Βάρος	1000			624			541			348			82			193		
	#	COR	CTR	#	COR	CTR	#	COR	CTR	#	COR	CTR	#	COR	CTR	#	COR	CTR
$\Delta 1$	190	0	0	179	75	37	167	270	158	161	517	166	260	385	233	178	16	13
$\Delta 2$	119	0	0	125	35	16	126	40	21	135	234	69	116	1	0	111	14	11
$\Delta 3$	110	0	0	87	670	216	89	427	160	105	24	5	75	167	65	58	603	329
$\Delta 4$	184	0	0	188	9	4	198	115	66	179	17	5	116	373	222	234	336	282
$\Delta 5$	186	0	0	199	123	68	201	110	71	203	179	63	191	2	1	196	12	11
$\Delta 6$	208	0	0	219	85	60	216	34	28	214	25	11	239	70	60	219	16	19

Εικόνα 2 από το λογισμικό MAD

Τα βήματα της μεθόδου CHC

Βήμα 1ο : Ο πίνακας δεδομένων $T(n,p)$ αναλύεται με την Παραγοντική Ανάλυση των Αντιστοιχιών

Βήμα 2ο : Με τη χρήση της μεθόδου **KARAP**, για κάθε «αντικείμενο» εντοπίζεται με φθίνουσα σειρά ο 1ος έως τον $p^{10\text{στο}}$ «**πλησιέστερο συγγενή**»

Βήμα 3ο : Για κάθε «αντικείμενο», αντιστοιχείται η **θέση της κατάταξης** κάθε «πλησιέστερου συγγενή» με συγκεκριμένο βαθμό ανάλογο του πλήθους των μεταβλητών.

Βήμα 4ο : Από το βήμα 3 βάσει του πλήθους των βαθμών που συγκέντρωσαν οι p μεταβλητές, εντοπίζεται η **ενιαία σημαντικότητα** της κάθε μεταβλητής,

Βήμα 5ο : Χρησιμοποιώντας την κατάταξη των μεταβλητών, προκύπτει με τη χρήση της Ιεραρχικής Ανάλυσης η **βαρύτητα** κάθε μεταβλητής.

Βήμα 6ο : Στη συνέχεια χρησιμοποιώντας τα στοιχεία του 5^{ου} βήματος σε συνδυασμό με τα στοιχεία του 1^{ου} βήματος, απορρέει η **φθίνουσα ταξινόμηση** των «αντικειμένων».

Παρατήρηση: Με την διαδικασία της Σύνθετης Ιεραρχικής ταξινόμησης τα «αντικείμενα» στο σύνολό τους αξιολογήθηκαν με την εκ των προτέρων ενιαία βαρύτητα των μεταβλητών και όχι ως υποσύνολα με μεταβλητές των οποίων η σημαντικότητα μεταβάλλεται από κλάση σε κλάση.

Εφαρμογή της μεθόδου CHC

Ο πίνακας 1 παρουσιάζει **τις απαντήσεις** 20 υποψήφιων εργαζόμενων (με βάση 5βάθμια κλίμακα Likert) σε 12 ερωτήματα στα οποία ανταποκρίθηκαν με σκοπό την επιλογή τους για πρόσληψη σε εταιρεία.

Πίνακας 1: Πίνακας δεδομένων

INT	EP1	EP2	EP3	EP4	EP5	EP6	EP7	EP8	EP9	EP10	EP11	EP12
I1	5	3	4	4	4	2	4	3	3	2	2	4
I2	2	3	5	3	4	3	3	4	3	4	3	5
I3	4	4	3	3	5	5	3	3	3	1	4	3
I4	3	3	3	3	5	2	4	3	3	4	4	3
I5	2	5	4	4	5	3	2	4	5	2	4	4
I6	2	2	5	5	2	5	4	4	5	4	3	5
I7	5	4	4	5	4	4	4	4	4	5	5	4
I8	4	4	5	3	3	5	4	5	3	4	4	5
I9	4	3	4	2	3	3	2	5	5	2	4	2
I10	4	3	4	3	3	5	3	4	3	5	4	4
I11	2	4	4	3	3	5	4	5	4	1	4	4
I12	5	4	2	4	3	4	4	4	4	5	5	4
I13	5	4	4	2	4	4	4	4	3	5	5	5
I14	2	2	5	5	3	4	4	4	4	5	4	2
I15	5	5	5	4	3	3	5	5	4	4	5	4
I16	5	5	5	2	5	4	3	2	3	2	5	2
I17	5	5	4	4	3	3	5	2	4	3	5	4
I18	5	2	5	3	3	2	5	4	2	2	2	2
I19	2	2	5	4	3	2	5	5	2	2	5	5
I20	2	3	4	5	2	4	4	4	2	3	2	5

Η ανάλυση του πίνακα δεδομένων πραγματοποιείται με την μέθοδο KARAP

Συμβολή της μεθόδου KARAP

Ο Δάσκαλος J.P. Benzecri θεωρούσε ότι σε μία ανάλυση AFC οι 7 πρώτοι παραγοντικοί άξονες αρκούν για την μελέτη ενός πίνακα T(n,p).

Για να γίνει κατανοητή η χρησιμότητα της μεθόδου KARAP ειδικά στην ερμηνεία του παραγοντικού επιπέδου 1x2 όταν το νέφος των μεταβλητών διαθέτει αρκετούς παραγοντικούς άξονες, όπου ελλοχεύουν παρερμηνείες για το ποια «αντικείμενα» συνδέονται με συγκεκριμένες μεταβλητές.

Ένα αριθμητικό παράδειγμα θα παρουσιάσει ανάγλυφα την χρησιμότητα της μεθόδου KARAP.

Κωδικοποιημένος πίνακας 3 μεταβλητών

INT	A	B	C
I1	1	4	3
I2	2	2	1
I3	1	4	2
I4	2	2	3
I5	1	1	4
I6	4	3	4
I7	2	4	3
I8	1	4	2
I9	3	2	4
I10	4	1	4



Ο λογικός πίνακας 0-1 των 3 μεταβλητών

INT	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
I1	1	0	0	0	0	0	0	1	0	0	1	0
I2	0	1	0	0	0	1	0	0	1	0	0	0
I3	1	0	0	0	0	0	0	1	0	1	0	0
I4	0	1	0	0	0	1	0	0	0	0	1	0
I5	1	0	0	0	1	0	0	0	0	0	0	1
I6	0	0	0	1	0	0	1	0	0	0	0	1
I7	0	1	0	0	0	0	0	1	0	0	1	0
I8	1	0	0	0	0	0	0	1	0	1	0	0
I9	0	0	1	0	0	1	0	0	0	0	0	1
I10	0	0	0	1	1	0	0	0	0	0	0	1

Χαρακτηριστικές ρίζες και ποσοστά ερμηνείας των παραγοντικών αξόνων

ΣΥΝΟΛΙΚΗ ΑΔΡΑΝΕΙΑ 3,00000			
ΑΞΟΝ	ΑΔΡΑΝΕΙΑ	%ΕΡΜΗΝΕΙΑΣ	ΑΘΡΟΙΣΜΑ ΙΣΤΟΓΡΑΜΜΑ ΧΑΡΑΚΤΗΡ ΡΙΖΩΝ
1	0,8583072	28,61	28,61 *****
2	0,7711879	25,71	54,32 *****
3	0,4511810	15,04	69,36 *****
4	0,3538555	11,80	81,15 *****

ΣΥΝΕΧΕΙΑ ΤΗΣ ΜΕΘΟΔΟΥ KARAP

Πίνακας 1

Οι θέσεις των μεταβλητών και των «αντικειμένων» στα 4 τεταρτημόρια του παραγοντικού επιπέδου 1x2

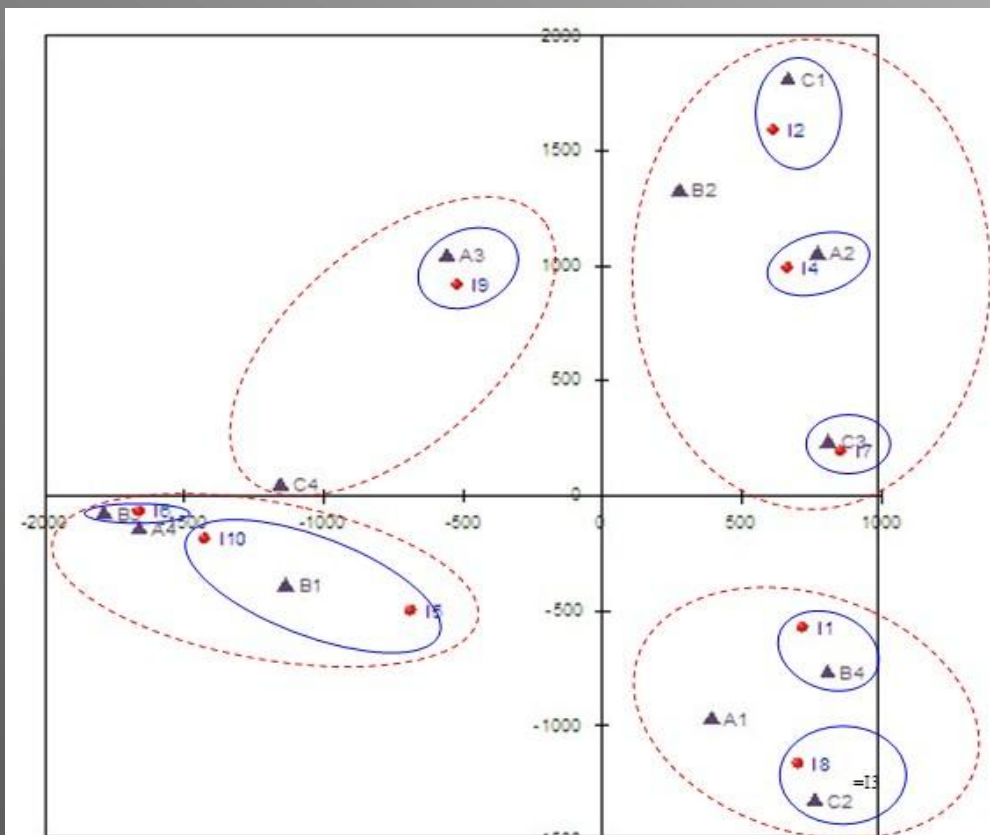
1ο	2ο	3ο	4ο		1ο	2ο	3ο	4ο
A2	A3	A4	A1		I2	I9	I5	I1
B2	C4	B1	B4		I4		I6	I3
C1		B3	C2		I7		I10	I8
C3								

Πίνακας 2

Οι θέσεις των μεταβλητών και των «αντικειμένων» με βάση τα αποτελέσματα της μεθόδου KARAP

A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
	I4	I9		I5		I6	I1	I2	I3	I7	
				I10					I8		

Η απεικόνιση του παραγοντικού επιπέδου 1x2 παρουσιάζεται στο παρακάτω σχήμα



A) Με συνεχή περιγράμματα περιγράφονται οι ομαδοποιήσεις με βάση την ανάλυση του λογικού πίνακα 0-1 με την μέθοδο KARAP, η οποία προσφέρει το 100% της πληροφορίας, αποδίνοντας την κύρια σύνδεση των «αντικειμένων» με τις μεταβλητές που συνδέονται **εντονότερα**, αφήνοντας κάποιες μεταβλητές που δεν διαδραματίζουν τόσο σημαντικό ρόλο στις αλληλεπιδράσεις των «αντικειμένων», όπως οι μεταβλητές A1, A4, B2 και C4 (Πίνακας 2)

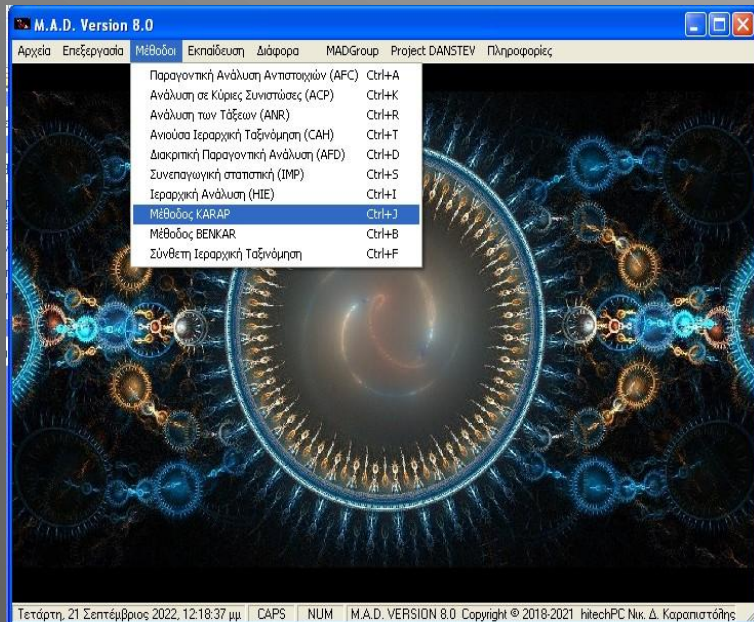
B) Με διακεκομμένα περιγράμματα περιγράφονται οι ομαδοποιήσεις με βάση την ερμηνεία του παραγοντικού επιπέδου 1x2, απ' όπου αντλείται το 54,32% της συνολικής πληροφορίας που προσφέρει ο πίνακας δεδομένων. (Πίνακας 1)

Η εικόνα είναι ενδεικτική για την υπεροχή της μεθόδου KARAP, όσον αφορά την ερμηνεία των σημείων ενός παραγοντικού επιπέδου 1x2 με καθορισμένο ποσοστό ερμηνείας

Συνέχεια της μεθόδου CHC με την εφαρμογή της μεθόδου KARAP

Αρχικά ο πίνακας δεδομένων με την χρήση του λογισμικού MAD, αναλύεται με την μέθοδο της Παραγοντικής Ανάλυσης των Αντιστοιχιών, ενώ στη συνέχεια εφαρμόζεται η μέθοδος KARAP, όπου εντοπίζονται ιεραρχικά για κάθε ερωτώμενο ο 1ος έως τον 12ο «πλησιέστερο συγγενή» του

Πίνακας 2: Πίνακας των πλησιέστερων συγγενών των 20 ερωτώμενων



ind	1ο	2ο	3ο	4ο	5ο	6ο	7ο	8ο	9ο	10ο	11ο	12ο
I1	EP7	EP1	EP3	EP5	EP4	EP12	EP2	EP8	EP9	EP11	EP6	EP10
I2	EP12	EP3	EP8	EP4	EP5	EP7	EP11	EP9	EP10	EP6	EP2	EP1
I3	EP5	EP2	EP6	EP11	EP9	EP1	EP3	EP8	EP7	EP12	EP4	EP10
I4	EP5	EP11	EP7	EP2	EP3	EP10	EP9	EP4	EP8	EP12	EP1	EP6
I5	EP9	EP2	EP5	EP11	EP8	EP3	EP4	EP12	EP6	EP7	EP1	EP10
I6	EP4	EP12	EP6	EP8	EP9	EP3	EP7	EP10	EP11	EP2	EP5	EP1
I7	EP11	EP4	EP7	EP9	EP10	EP2	EP3	EP8	EP6	EP1	EP12	EP5
I8	EP8	EP6	EP12	EP3	EP7	EP11	EP2	EP9	EP4	EP10	EP5	EP1
I9	EP9	EP8	EP11	EP3	EP2	EP5	EP6	EP1	EP7	EP4	EP12	EP10
I10	EP6	EP10	EP11	EP8	EP12	EP3	EP9	EP7	EP2	EP4	EP1	EP5
I11	EP8	EP6	EP9	EP3	EP2	EP11	EP12	EP7	EP4	EP5	EP1	EP10
I12	EP11	EP10	EP2	EP9	EP7	EP1	EP6	EP4	EP12	EP8	EP5	EP3
I13	EP11	EP2	EP12	EP1	EP10	EP7	EP6	EP5	EP8	EP3	EP9	EP4
I14	EP4	EP3	EP10	EP9	EP8	EP7	EP6	EP11	EP12	EP5	EP2	EP1
I15	EP7	EP11	EP3	EP2	EP8	EP9	EP1	EP4	EP12	EP5	EP6	EP10
I16	EP2	EP5	EP11	EP1	EP3	EP6	EP9	EP7	EP8	EP12	EP4	EP10
I17	EP2	EP11	EP7	EP1	EP9	EP4	EP3	EP5	EP12	EP6	EP8	EP10
I18	EP7	EP3	EP1	EP8	EP4	EP5	EP11	EP2	EP12	EP9	EP6	EP10
I19	EP7	EP8	EP3	EP12	EP4	EP11	EP5	EP9	EP2	EP6	EP10	EP1
I20	EP4	EP12	EP7	EP8	EP3	EP6	EP10	EP9	EP2	EP11	EP5	EP1

Αποτελέσματα της μεθόδου KARAP

Πίνακας 3: Αντιστοίχιση της θέσης κατάταξης της μεταβλητής με βαθμό

ind	EP1	EP2	EP3	EP4	EP5	EP6	EP7	EP8	EP9	EP10	EP11	EP12
I1	11	6	10	8	9	2	12	5	4	1	3	7
I2	1	2	11	9	8	3	7	10	5	4	6	12
I3	7	11	6	2	12	10	4	5	8	1	9	3
I4	2	9	8	5	12	1	10	4	6	7	11	3
I5	2	11	7	6	10	4	3	8	12	1	9	5
I6	1	3	7	12	2	10	6	9	8	5	4	11
I7	3	7	6	11	1	4	10	5	9	8	12	2
I8	1	6	9	4	2	11	8	12	5	3	7	10
I9	5	8	9	3	7	6	4	11	12	1	10	2
I10	2	4	7	3	1	12	5	9	6	11	10	8
I11	2	8	9	4	3	11	5	12	10	1	7	6
I12	7	10	1	5	2	6	8	3	9	11	12	4
I13	9	11	3	1	5	6	7	4	2	8	12	10
I14	1	2	11	12	3	6	7	8	9	10	5	4
I15	6	9	10	5	3	2	12	8	7	1	11	4
I16	9	12	8	2	11	7	5	4	6	1	10	3
I17	9	12	6	7	5	3	10	2	8	1	11	4
I18	10	5	11	8	7	2	12	9	3	1	6	4
I19	1	4	10	8	6	3	12	11	5	2	7	9
I20	1	4	8	12	2	7	10	9	5	6	3	11
sum	90	144	157	127	111	116	157	148	139	84	165	122

Πίνακας 4: Ο «πλησιέστερος συγγενής» για κάθε ένα από τους 20 ερωτώμενους

IND	EP1	EP2	EP3	EP4	EP5	EP6	EP7	EP8	EP9	EP10	EP11	EP12
ΠΛΗΘΟΣ	0	2	0	3	2	1	4	2	2	0	3	1
1		I16		I6	I3	I10	I1	I8	I5		I7	I2
2		I17		I14	I4		I15	I11	I9		I12	
3				I20			I18				I13	
4							I19					

Αποτελέσματα της μεθόδου CHC

Σημαντικότητα-Βαρύτητα των ερωτήσεων

Στη συνέχεια χρησιμοποιώντας την βαθμολόγηση των ερωτημάτων (πίνακας 3), προκύπτει η **βαρύτητα** κάθε ερώτησης. Το αποτέλεσμα παρουσιάζεται στον πίνακα 5

Πίνακας 5 Σημαντικότητα-Βαρύτητα των ερωτήσεων

ΚΑΤΑΤΑΞΗ	ΕΡΩΤΗΣΗ	ΒΑΘΜΟΙ	ΒΑΡΥΤΗΤΑ
1η	EP11	165	10.58
2η	EP7	157	10.06
3η	EP3	157	10.06
4η	EP8	148	9.49
5η	EP2	144	9.23
6η	EP9	139	8.91
7η	EP4	127	8.14
8η	EP12	122	7.82
9η	EP6	116	7.44
10η	EP5	111	7.12
11η	EP1	90	5.77
12η	EP10	84	5.38

Αποτελέσματα της μεθόδου CHC (Συνέχεια)

Ταξινόμηση των 20 «αντικειμένων»

Ακολουθεί η ταξινόμηση των «αντικειμένων» κάνοντας χρήση των στοιχείων του πίνακα 3 και του πίνακα 5. Το αποτέλεσμα παρουσιάζεται στον πίνακα 6.

Πίνακας 6: Πίνακας ταξινόμησης των 20 «αντικειμένων»

ΚΑΤΑΤΑΞΗ	ΑΤΟΜΟ	EP1	EP2	EP3	EP4	EP5	EP6	EP7	EP8	EP9	EP10	EP11	EP12	ΒΑΘΜΟΙ
1ος	I15	5	3	4	4	4	2	4	3	3	2	2	4	707,28
2ος	I19	2	3	5	3	4	3	3	4	3	4	3	5	698,31
3ος	I11	4	4	3	3	5	5	3	3	3	1	4	3	691,32
4ος	I9	3	3	3	3	5	2	4	3	3	4	4	3	690,5
5ος	I8	2	5	4	4	5	3	2	4	5	2	4	4	687,64
6ος	I5	2	2	5	5	2	5	4	4	5	4	3	5	686,01
7ος	I7	5	4	4	5	4	4	4	4	4	5	5	4	682,58
8ος	I17	4	4	5	3	3	5	4	5	3	4	4	5	681,85
9ος	I4	4	3	4	2	3	3	2	5	5	2	4	2	678,19
10ος	I18	4	3	4	3	3	5	3	4	3	5	4	4	677,35
11ος	I2	2	4	4	3	3	5	4	5	4	1	4	4	676,14
12ος	I20	5	4	2	4	3	4	4	4	4	5	5	4	667,77
13ος	I16	5	4	4	2	4	4	4	4	3	5	5	5	666,21
14ος	I14	2	2	5	5	3	4	4	4	4	5	4	2	663,08
15ος	I6	5	5	5	4	3	3	5	5	4	4	5	4	662,49
16ος	I3	5	5	5	2	5	4	3	2	3	2	5	2	661,43
17ος	I1	5	5	4	4	3	3	5	2	4	3	5	4	659,2
18ος	I10	5	2	5	3	3	2	5	4	2	2	2	2	656,41
19ος	I12	2	2	5	4	3	2	5	5	2	2	5	5	648,89
20ος	I13	2	3	4	5	2	4	4	4	2	3	2	5	646,42

ΣΥΜΠΕΡΑΣΜΑ

1ο) Η μέθοδος της Σύνθετης Ιεραρχικής Ταξινόμησης προτείνει μία νέα μέθοδο ταξινόμησης, όπου τα κριτήρια διαμορφώνονται ενιαία από τα ίδια στοιχεία που αξιολογούνται και όχι από αυθαίρετους υπολογισμούς ή a priori υποθέσεις.

2ο) Είναι φανερό ότι η μέθοδος CHC και η μέθοδος VACOR χρησιμοποιούν δύο διαφορετικούς όρους, η μεν μέθοδος CHC τον όρο «**πλησιέστερο συγγενή**», η δε VACOR «**τον πλησιέστερο γείτονα**», με αποτέλεσμα να διαφέρουν ουσιαστικά στη φιλοσοφία προσέγγισης του μέτρου «**απόσταση**», η οποία καθορίζει την ταξινόμησή τους, ανάλογα με το πλήθος των χρησιμοποιούμενων p κριτηρίων.

3ο) Στη μέθοδο VACOR οι μεταβλητές δεν επηρεάζουν **ενιαία** τη διαμόρφωση των κλάσεων της ταξινόμησης, αλλά διαφορετικά τη κάθε κλάση, ενώ αντιθέτως η προτεινόμενη μέθοδος CHC αξιολογεί **ενιαία τη σημαντικότητα των μεταβλητών στο σύνολο των κλάσεων** της διαμορφούμενης τυπολογίας της ιεραρχίας (πίνακας 5),

4ο) Η ανάλυση των δεδομένων καταλήγει σε μία **φθίνουσα ιεραρχική ταξινόμηση** των «αντικειμένων» (πίνακας 6), η οποία στη συνέχεια μπορεί να χρησιμοποιηθεί ως διαμελισμός του συνόλου των «αντικειμένων», σε όσες κλάσεις κριθεί σκόπιμο από τον αναλυτή, ανάλογη διαδικασία με εκείνη που ισχύει με την τομή του δένδρογράμματος στη μέθοδο της Ανιούσας Ιεραρχικής Ταξινόμησης.

11 0 ΠΑΝΕΛΛΗΝΙΟ ΣΥΝΕΔΡΙΟ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ
ΓΡΕΒΕΝΑ 29 Σεπτεμβρίου - 2 Οκτωβρίου 2022

Αξιολόγηση των κλάσεων της Ανιούσας Ιεραρχικής Ταξινόμησης

Δρ. Δημήτριος Καραπιστόλης

ΕΙΣΑΓΩΓΗ ΣΤΗ ΜΕΘΟΔΟ

Με την μέθοδο ταξινόμησης που προτείνεται από τον Δρα Δημήτριο Καραπιστόλη η οποία καλείται «**Σύνθετη Ιεραρχική Ταξινόμηση**» (**Complex Hierarchical Classification-CHC-**) και τον καθορισμό από αυτή του νέου όρου «**πλησιέστερος συγγενής**» που χρησιμοποιεί, αποδείχθηκε ότι ο χαρακτηρισμός (**σημαντικότητα**) κάθε μεταβλητής που αφορά μία συγκεκριμένη κλάση είναι διαφορετικός σε κάθε κλάση της Ανιούσας Ιεραρχικής Ταξινόμησης με βάση τον αλγόριθμο του Ward (-μέθοδος VACOR-).

Συνεπώς η προσδιοριζόμενη από τη μέθοδο VACOR σημαντικότητα μιας μεταβλητής X_k **ΔΕΝ** αποτελεί ένα **ενιαίο χαρακτηρισμό** για το σύνολο των κλάσεων της ταξινόμησης, αλλά αφορά κάθε κλάση χωριστά και για όσα στοιχεία συμμετέχουν σε κάθε κλάση.

Εύλογα, λοιπόν, ανακύπτει το ερώτημα: «με βάση την διαφορετική σημαντικότητα των μεταβλητών που παρατηρείται σε κάθε κλάση της Ανιούσας Ιεραρχίας, ποια κλάση παρουσιάζει για τα στοιχεία που συμμετέχουν σε αυτή, την **ισχυρότερη παρουσία** στη διαμορφούμενη τυπολογία».

Ακριβώς σε αυτό το ερώτημα θα δοθεί η απάντηση, αφού πρώτα αξιολογηθεί χωριστά η κάθε κλάση της διαμορφούμενης τυπολογίας της Ανιούσας Ιεραρχίας με την μέθοδο της Σύνθετης Ιεραρχικής Ταξινόμησης (μέθοδος CHC).

Η σημαντικότητα των μεταβλητών στη μέθοδο VACOR

Θεωρούμε ένα πίνακα με 84 γραμμές και 6 μεταβλητές Δ_1 έως Δ_6 . Ο χώρος των κόμβων που δημιουργείται είναι ο R^6 , ενώ δημιουργούνται $2 \cdot 84 - 1 = 167$ κλάσεις. Η 1η διάσπαση της κλάσης 167 δημιουργεί το δίπολο των κλάσεων 166-161, ενώ η αμέσως επόμενη διάσπαση της κλάσης 166, δημιουργεί το δίπολο 165-163

d84.cla																		
Συμβολή των μεταβλητών στον χαρακτηρισμό των κόμβων																		
Κόμβος	167			166			165			164			163			162		
A(l)	166			165			164			159			156			157		
B(l)	161			163			162			148			155			160		
Βάρος	1000			624			541			348			82			193		
	#	COR	CTR	#	COR	CTR	#	COR	CTR	#	COR	CTR	#	COR	CTR	#	COR	CTR
Δ_1	190	0	0	179	75	37	167	270	158	161	517	166	260	385	233	178	16	13
Δ_2	119	0	0	125	35	16	126	40	21	135	234	69	116	1	0	111	14	11
Δ_3	110	0	0	87	670	216	89	427	160	105	24	5	75	167	65	58	603	329
Δ_4	184	0	0	188	9	4	198	115	66	179	17	5	116	373	222	234	336	282
Δ_5	186	0	0	199	123	68	201	110	71	203	179	63	191	2	1	196	12	11
Δ_6	208	0	0	219	85	60	216	34	28	214	25	11	239	70	60	219	16	19

Από τον παραπάνω πίνακα η στήλη $COR(X_k)$ θεωρείται ως η **σχετική συμβολή** της μεταβλητής X_k στην απόκλιση της κλάσης K_i από το κέντρο του νέφους $K_{(2n-1)}$, οπότε σχετικά με την μεταβλητή Δ_1 έχουμε $COR_{166}(\Delta_1)=75$, που σημαίνει ότι η μεταβλητή Δ_1 συμβάλει στη **ΑΠΟΚΛΙΣΗ** της κλάσης 166 από τον κόμβο 167 σε ποσοστό 75%, το οποίο ερμηνεύεται και ως το ποσοστό που **χαρακτηρίζει** η μεταβλητή Δ_1 την κλάση 166 στη διάσπαση της κλάσης 167.

Από τον παραπάνω πίνακα προκύπτει ότι ο χαρακτηρισμός της Δ_1 (όπως και των άλλων μεταβλητών) που αφορά την κλάση 166 είναι **διαφορετικός** σε κάθε κλάση της ιεραρχίας. Συνεπώς επιβεβαιώνεται ότι ο χαρακτηρισμός των μεταβλητών $\Delta_1, \Delta_2, \Delta_3, \Delta_4, \Delta_5, \Delta_6$ **ΔΕΝ** αποτελεί ένα **ενιαίο χαρακτηρισμό** για το σύνολο των κλάσεων της ταξινόμησης, αλλά αφορά μόνο τη κάθε κλάση χωριστά με τα «αντικείμενα» που συμμετέχουν σε αυτή.

Τα βήματα της αξιολόγησης

Η διαδικασία της αξιολόγησης που προβλέπεται από το λογισμικό MAD είναι η ακόλουθη :

1^ο βήμα: Εντοπισμός των «αντικειμένων» στις διαμορφούμενες κλάσεις που θέλουμε να αξιολογήσουμε, αφού προηγουμένως τις εντοπίσουμε κατά την φάση της ανάλυσης με την μέθοδο VACOR

2^ο βήμα:Εξάγουμε το αρχείο που προσδιορίζει με βάση την κατανομή z , ποιες μεταβλητές είναι οι σημαντικότερες για κάθε ένα «αντικείμενο» που περιλαμβάνεται στο δείγμα.

3^ο βήμα : Διαμόρφωση για κάθε αξιολογούμενη κλάση ενός πίνακα με τις τιμές της κατανομής z που αντιστοιχούν στα «αντικείμενα» που συμμετέχουν σε κάθε κλάση.

4^ο βήμα:Αξιολόγηση κάθε κλάσης χωριστά με τη μέθοδο CHC

5^ο βήμα: Δημιουργία συγκριτικού πίνακα αξιολόγησης των κλάσεων

6^ο βήμα: Εντοπισμός της κλάσης με την **ισχυρότερη παρουσία** της συγκεκριμένης τυπολογίας, με βάση την ενιαία σημαντικότητα των μεταβλητών.

Εφαρμογή της μεθόδου

Εφαρμογή με ένα παράδειγμα στο αρχείο om1_d49.cla που αφορά 99 «αντικείμενα» τα οποία ταξινομήθηκαν σε πέντε κλάσεις με την μέθοδο VACOR

1) Εντοπισμός των «αντικειμένων» στις διαμορφούμενες 5 κλάσεις που θέλουμε να αξιολογήσουμε

Τυπολογία της Ιεραρχίας
σε 5 κλάσεις 180,186,191,192,193

Προβολή των αποτελεσμάτων της ταξινόμησης VACOR

om1_d49.cla

ΟΙ 5 ΚΟΜΒΟΙ ΤΗΣ ΑΠΟΥΣΙΑΣ ΙΕΡΑΡΧΙΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ

Κομβός	A(i)	B(i)	Βάρος	IS	IS	IS	IS
193	190	182	0,38297	0,0047	0,04409	0,04738	0,51795
194	193	192	0,5739	0,0059	0,04999	0,04149	0,45354
195	186	180	0,36862	0,00865	0,05864	0,03284	0,35898
196	194	195	0,94852	0,01462	0,07325	0,01822	0,19916
197	196	191	1	0,0182	0,09148	0	0

ΕΝΤΟΠΙΣΜΟΣ ΤΩΝ 4 ΔΙΠΛΩΝ ΤΗΣ ΙΕΡΑΡΧΙΑΣ

Δίπλο	Is	Is
194	193	192
195	186	180
196	194	195
197	196	191

ΟΙ 5 ΚΛΑΣΕΙΣ

Κλάση	Κομβοί
180	191
186	195
191	197
192	193
193	194

Πλήθος κλάσεων = 197

Προβολή των αποτελεσμάτων της ταξινόμησης VACOR

om1_d49.cla

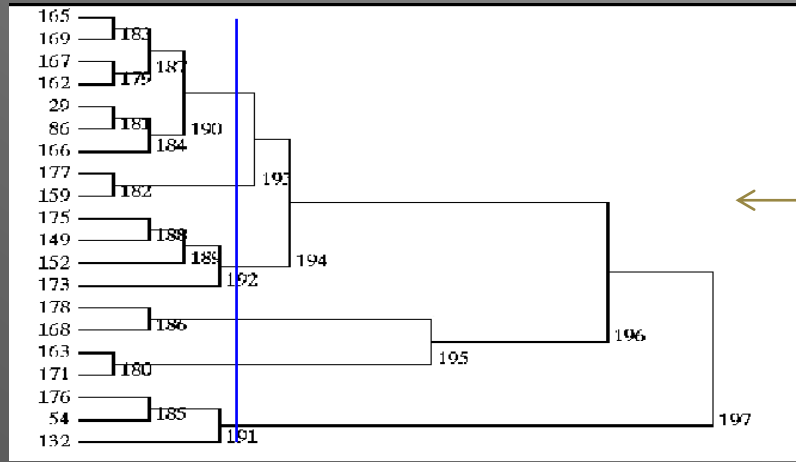
ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΚΛΑΣΕΩΝ ΤΗΣ ΑΠΟΥΣΙΑΣ ΙΕΡΑΡΧΙΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ

Κομβός	180	181	182	183	184	185	186	187
A(i)	171	29	177	164	181	176	178	183
B(i)	170	86	160	168	165	54	167	175
Πλήθος	7	2	7	16	5	5	26	26
101	452	209	11	452	389	60	11	
172	1903	703	246	1903	890	315	246	
241		643	355	954	1055	99	395	
1481		705	368	759	1305	536	368	
845		1433	1307	1548	864	1546	1307	
502		645	213			567	213	
1623		1105	1383			914	1383	
			312			702	312	
			745			1020	745	
			922			298	922	
			882			1019	882	
			73			1626	73	
			88			1127	88	
			693			1200	693	
			644			301	644	

ΠΙΝΑΚΑΣ ΣΥΜΠΤΥΞΗΣ-ΚΩΔΙΚΟΙ

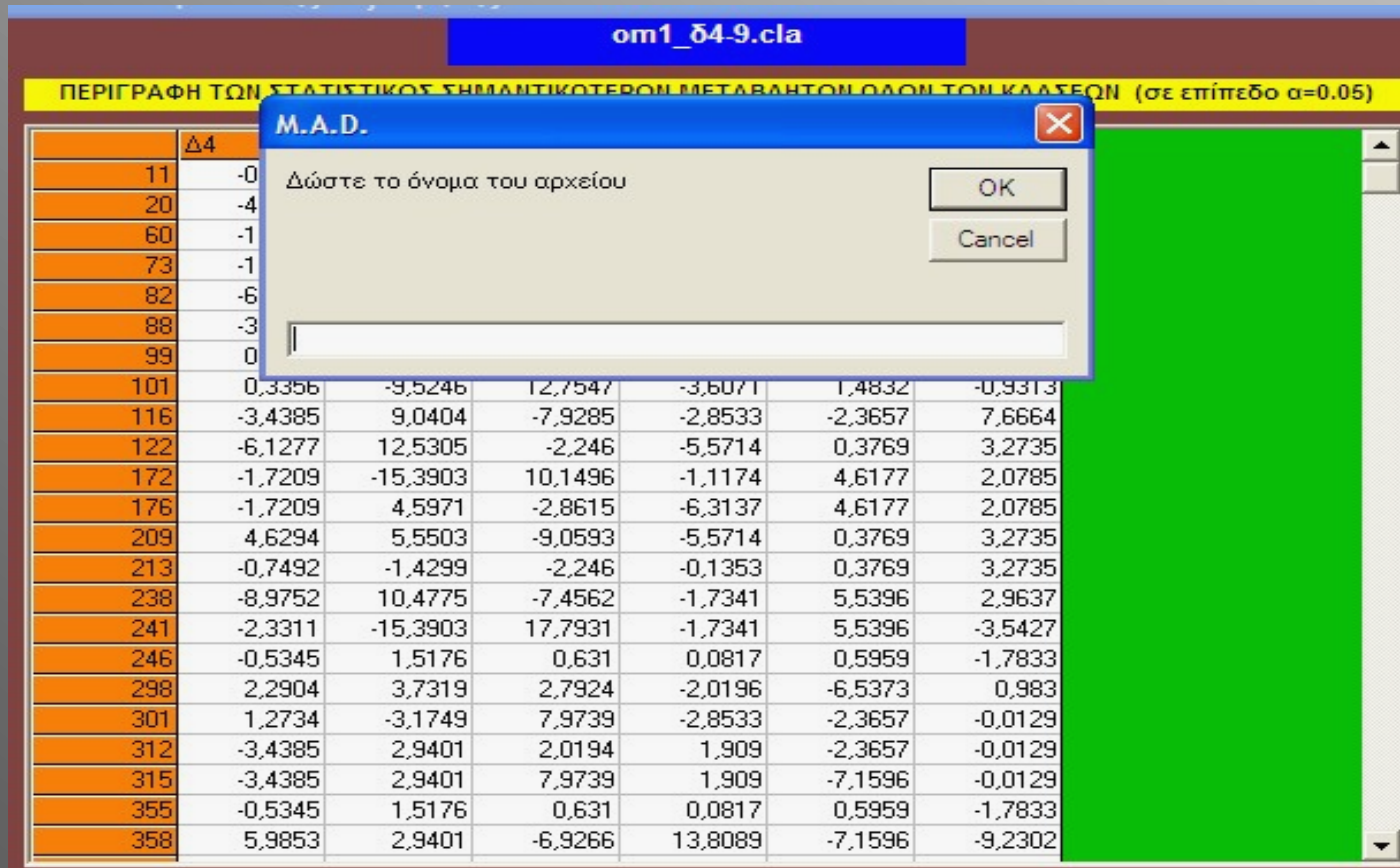
Κλάσεις	180	186	191
1	101	60	389
2	172	315	890
3	241	99	1055
4	1481	536	1305
5	845	1546	864
6	502	567	1185
7	1623	914	1321
8		702	
9		1020	
10		298	
11		1019	
12		1626	
13		1127	
14		1200	
15		301	
16		1269	
17		399	
18		1172	
19		1114	
20		1156	
21		1530	
22		1540	
23		1580	
24		1619	
25		1570	
26		1633	
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			

ΚΛΑΣΕΙΣ	180	186	191	192	193
1	101	60	389	116	11
2	172	315	890	704	246
3	845	99	1055	879	355
4	241	536	1305	1086	368
5	1481	1546	864	378	1307
6	502	567	1185	596	213
7	1623	914	1321	678	1383
8		702		641	312
9		1020		1696	745
10		298		629	922
11		1019		1279	882
12		1626		358	73
13		1127		1712	88
14		1200		1328	693
15		301		122	644
16		1269		630	1006
17		399		635	20
18		1172		1085	176
19		1114		238	1482
20		1156		420	990
21		1530		820	1000
22		1540			82
23		1580			1140
24		1619			553
25		1570			555
26		1633			1694
27					452
28					1503
29					554
30					759
31					1548
32					209
33					703
34					643
35					705
36					1433
37					645
38					1105



Εφαρμογή της μεθόδου (Συνέχεια)

2) Εξαγωγή του αρχείου om1_d49_z.cla που προσδιορίζει με βάση την κατανομή z, ποιες μεταβλητές είναι οι σημαντικότερες για κάθε ένα «αντικείμενο» που περιλαμβάνεται στην ανάλυση.



The screenshot shows a software window titled "om1_δ4-9.cla". Below the title bar is a yellow header with the text "ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΣ ΣΗΜΑΝΤΙΚΟΤΕΡΩΝ ΜΕΤΑΒΛΗΤΩΝ ΟΔΩΝ ΤΩΝ ΚΑΔΕΣΩΝ (σε επίπεδο $\alpha=0.05$)". The main area contains a table with a column labeled "Δ4" and several rows of numerical data. A dialog box titled "M.A.D." is overlaid on the table, with the text "Δώστε το όνομα του αρχείου" and "OK" and "Cancel" buttons.

	Δ4					
11	-0					
20	-4					
60	-1					
73	-1					
82	-6					
88	-3					
99	0					
101	0,3356	-9,5246	12,7547	-3,6071	1,4832	-0,9313
116	-3,4385	9,0404	-7,9285	-2,8533	-2,3657	7,6664
122	-6,1277	12,5305	-2,246	-5,5714	0,3769	3,2735
172	-1,7209	-15,3903	10,1496	-1,1174	4,6177	2,0785
176	-1,7209	4,5971	-2,8615	-6,3137	4,6177	2,0785
209	4,6294	5,5503	-9,0593	-5,5714	0,3769	3,2735
213	-0,7492	-1,4299	-2,246	-0,1353	0,3769	3,2735
238	-8,9752	10,4775	-7,4562	-1,7341	5,5396	2,9637
241	-2,3311	-15,3903	17,7931	-1,7341	5,5396	-3,5427
246	-0,5345	1,5176	0,631	0,0817	0,5959	-1,7833
298	2,2904	3,7319	2,7924	-2,0196	-6,5373	0,983
301	1,2734	-3,1749	7,9739	-2,8533	-2,3657	-0,0129
312	-3,4385	2,9401	2,0194	1,909	-2,3657	-0,0129
315	-3,4385	2,9401	7,9739	1,909	-7,1596	-0,0129
355	-0,5345	1,5176	0,631	0,0817	0,5959	-1,7833
358	5,9853	2,9401	-6,9266	13,8089	-7,1596	-9,2302

Εφαρμογή της μεθόδου (Συνέχεια)

3) Διαμόρφωση για κάθε αξιολογούμενη κλάση του πίνακα με τις τιμές της κατανομής z που αντιστοιχούν στα «αντικείμενα» που συμμετέχουν σε κάθε κλάση.

Διαμόρφωση κλάσεων για αξιολόγηση

Πίνακας δεδομένων om1_64-9_z.afc

INT	Δ4	Δ5	Δ6	Δ7	Δ1
11	-0,5345	1,5176	0,631	0,0817	
20	-4,8848	7,1634	-4,8654	-4,3151	
60	-1,7209	-2,0604	3,6369	4,0674	
73	-1,0769	2,9401	-6,9266	6,6712	

Πίνακας δεδομένων om1_64-9_5.afc

INT	180	186	191
I1	101	60	389
I2	172	315	890
I3	845	99	1055
I4	241	536	1305

Αρχεία κλάσεων

INT	EP1	EP2
101	0,3356	-9,5246
172	-1,7209	-15,3903
845	-3,4385	-15,3903
241	-2,3311	-15,3903
1481	-1,7209	-15,3903

INT	EP1	EP2
60	-1,7209	-2,0604
315	-3,4385	2,9401
99	0,3356	-0,7209
536	4,6294	-1,4294
1546	3,409	-2,0604

INT	EP1	EP2	E
389	5,9853	-15,3903	
890	5,9853	-15,3903	
1055	18,0871	-15,3903	
1305	10,9684	-15,3903	
864	15,3978	-15,3903	

INT	EP1	E
116	-3,4385	
704	-4,4328	
879	-1,0769	
1086	-4,4328	

INT	EP1	E
11	-0,5345	
246	-0,5345	
355	-0,5345	
368	-4,4328	

Ανοιγμα Αρχείων

ΕΠΙΣΤΡΟΦΗ

ΕΚΤΥΠΩΣΗ των αρχείων σε .afc και XLS

ΤΕΛΟΣ 1ης Φάσης

Εφαρμογή της μεθόδου CHC σε κάθε κλάση

Αρχικά αξιολογείται με τη μέθοδο CHC το βασικό αρχείο om1_d49_z.cla που προέκυψε από το 2^ο βήμα. Εμφανίζεται η αξιολόγηση των έξι μεταβλητών με βάση το σύνολο των 99 «αντικειμένων» του δείγματος, δίνοντας ένα νέο όνομα το om1_δ4-9_z_1.afc, το οποίο αποτελεί το ΝΕΟ βασικό αρχείο σύγκρισης για την επόμενη επιλογή «Συγκριτικός πίνακας αξιολόγησης των κλάσεων»

Πίνακας αξιολόγησης των έξι μεταβλητών με βάση το σύνολο των «αντικειμένων» του δείγματος

ΚΑΤΑΤΑΞΗ	ΕΡΩΤΗΣΗ	ΒΑΘΜΟΙ	ΒΑΡΥΤΗΤΑ
1η	EP2	412	19,82
2η	EP5	376	18,09
3η	EP3	344	16,55
4η	EP4	337	16,21
5η	EP1	312	15,01
6η	EP6	298	14,33

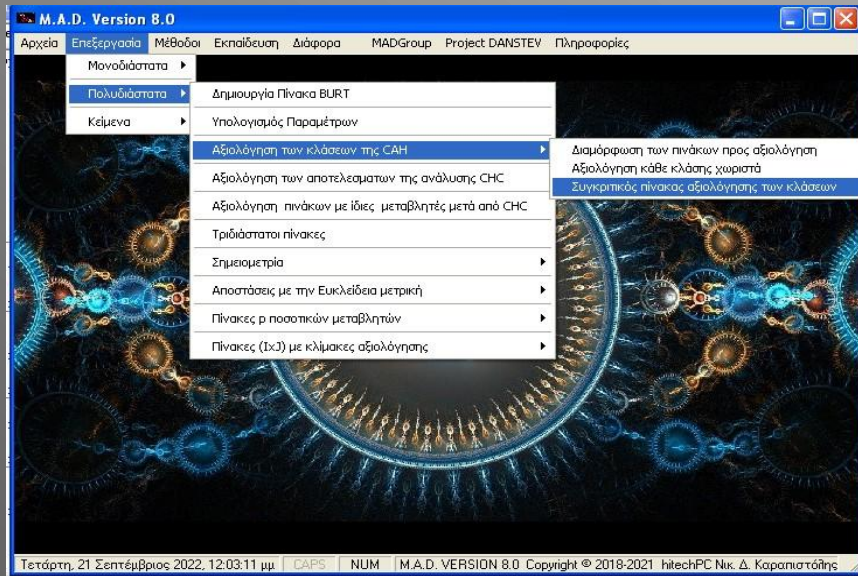
Πίνακας αξιολόγησης με βάση την μέθοδο CHC των έξι μεταβλητών σε κάθε κλάση της τυπολογίας

ΚΛΑΣΗ 180			
ΚΑΤΑΤΑΞΗ	ΕΡΩΤΗΣΗ	ΒΑΘΜΟΙ	ΒΑΡΥΤΗΤΑ
1η	EP3	41	27,89
2η	EP5	32	21,77
3η	EP4	28	19,05
4η	EP1	21	14,29
5η	EP6	18	12,24
6η	EP2	7	4,76
ΚΛΑΣΗ 186			
1η	EP3	139	25,46
2η	EP2	118	21,61
3η	EP1	97	17,76
4η	EP4	75	13,74
5η	EP5	61	11,17
6η	EP6	56	10,26
ΚΛΑΣΗ 191			
1η	EP4	37	25,17
2η	EP1	34	23,13
3η	EP5	26	17,69
4η	EP6	23	15,65
5η	EP2	17	11,56
6η	EP3	10	6,8
ΚΛΑΣΗ 192			
1η	EP2	108	24,49
2η	EP4	85	19,27
3η	EP6	85	19,27
4η	EP5	81	18,37
5η	EP1	45	10,2
6η	EP3	37	8,39
ΚΛΑΣΗ 193			
1η	EP5	174	22,39
2η	EP2	158	20,33
3η	EP3	114	14,67
4η	EP4	111	14,29
5η	EP6	111	14,29
6η	EP1	109	14,03

Αξιολόγηση των 5 κλάσεων

Στο παρακάτω πίνακα με την χρήση του λογισμικού MAD και με την επιλογή «Συγκριτικός πίνακας αξιολόγησης των κλάσεων» παρουσιάζεται η ταξινόμηση των 5 κλάσεων, εντοπίζοντας την κλάση με την **ισχυρότερη παρουσία των έξι μεταβλητών στη διαμόρφωση της συγκεκριμένης τυπολογίας**, η οποία δεν είναι άλλη από την κλάση 192 με την υψηλότερη βαθμολογία.

Πίνακας αξιολόγησης των 5 κλάσεων



INT	ΚΛΑΣΕΙΣ	ΒΑΘΜΟΙ	ΑΞΙΟΛΟΓΗΣΗ
I1	180	1648,43	DOWN
I2	186	1688,07	UP
I3	191	1641,12	DOWN
I4	192	1698,17	UP
I5	193	1695,93	UP
	ΟΛΕΣ ΜΑΖΙ	1687,40	

Με αντικειμενική διαδικασία και την χρήση μεθόδων της Ανάλυσης Δεδομένων, η προτεινόμενη μεθοδολογία εντοπίζει **«με βάση την διαφορετική σημαντικότητα των μεταβλητών»** που παρατηρείται σε κάθε κλάση της Ανιούσας Ιεραρχίας που σχηματίζεται με την μέθοδο VACOR, ποια κλάση παρουσιάζει για τα στοιχεία που συμμετέχουν σε αυτή, την **ισχυρότερη παρουσία** στη διαμορφούμενη τυπολογία.

Εξαγωγή τριδιάστατου πίνακα δεδομένων από αρχείο T(I,J) με $i=1,..n$ και $j=1,..p$)

Η διαδικασία αυτή είναι πρωτότυπη, δημιουργία του Δρος Δημητρίου Καραπιστόλη. Η αναγκαιότητα της δημιουργίας αυτής της επεξεργασίας οφείλεται στο γεγονός ότι πολλές φορές θέλουμε να εντοπίσουμε από ένα αρχείο p μεταβλητών, την κατανομή ΤΡΙΩΝ συγκεκριμένων μεταβλητών, ώστε να δημιουργηθεί ένας τριδιάστατος πίνακας, ο οποίος θα αναλυθεί με την διαδικασία που προβλέπεται από τον καθηγητή Δρ, Jean Paul Benzecri, η οποία πραγματοποιείται με το λογισμικό MAD, ανατρέχοντας στην διαδικασία «Επεξεργασία/ Πολυδιάστατα/Τριδιάστατοι πίνακες».

Αρχικά από τον πίνακα δεδομένων ο οποίος αποτελείται από τιμές μεταβλητών που αναφέρονται σε p διάφορα ερωτήματα, επιλέγονται και εξάγονται σε χωριστό αρχείο, οι ΤΡΕΙΣ μεταβλητές, τις οποίες θέλουμε να διασταυρώσουμε τις τιμές μεταξύ τους.

IND	A1	A2	A3
I1	2	3	3
I2	1	5	1
I3	1	1	2
I4	2	2	3
I5	1	3	1

Η 1η που θα επιλεγεί θεωρείται ως ΒΑΣΙΚΗ και επιβάλλεται να έχει ΔΥΟ ή ΤΡΕΙΣ διαβαθμίσεις, ενώ οι δύο άλλες θεωρούνται ως ΔΕΥΤΕΡΕΥΟΥΣΕΣ και μπορεί να έχουν δύο ή και περισσότερες διαβαθμίσεις. Αρχικά φορτώνουμε το αρχείο με τις τρεις μεταβλητές και δημιουργούμε τον αντίστοιχο λογικό πίνακα 0-1, το οποίο σώζουμε ως αρχείο με κατάληξη .afc. Φορτώνουμε στη συνέχεια το αρχείο λ.χ το ka_0-1.afc και στο ερώτημα «Πόσες διαβαθμίσεις έχει η βασική μεταβλητή;» θέτουμε στην προκειμένη περίπτωση το 2, επειδή διαλέξαμε ως βασική μεταβλητή την A1 η οποία έχει δύο διαβαθμίσεις.

Ακολουθούν δύο ερωτήματα για το πόσες διαβαθμίσεις έχουν οι δύο άλλες δευτερεύουσες μεταβλητές. Στο παράδειγμα δίνουμε τις τιμές 5 και 3, αφού οι μεταβλητές που ακολουθούν A2 και A3 παρουσιάζουν 5 και 3 διαβαθμίσεις αντίστοιχα. Αμέσως στην οθόνη παρουσιάζεται ο τριδιάστατος πίνακας έτοιμος με κάποιες μικρές διορθώσεις που θα γίνουν μετά την εκτύπωση του πίνακα αυτού, για επεξεργασία ως τριδιάστατος πλέον πίνακας. (βλέπε Οδηγίες χρήσης Πολυδιάστατα 2.2.6.4 Η επιλογή «Επεξεργασία πίνακα τριπλής εισόδου»

A1									
A3/A11					A3/A12				
A11	A31	A32	A33	ΣΥΝΟΛΟ	A12	A31	A32	A33	ΣΥΝΟΛΟ
A21	2	2	0	4	A21	2	4	0	6
A22	5	15	7	27	A22	7	13	4	24
A23	2	8	4	14	A23	5	11	7	23
A24	3	5	7	15	A24	3	8	6	17
A25	2	1	6	9	A25	1	0	0	1
ΑΘΡΟΙΣΜΑ	14	31	24	69	ΑΘΡΟΙΣΜΑ	18	36	17	71

69+71=140

Όμιλος λογισμικών MADGROUP



Στον όμιλο MADGROUP, όπου δεσπόζει το πρωτότυπο Ελληνικό λογισμικό MAD το οποίο «τρέχει» σε Windows XP, win10 & win11, συνδέονται τα παρακάτω προγράμματα:

A) Το λογισμικό **MAD**

Λογισμικό Ανάλυσης Δεδομένων



B) Το λογισμικό **DAM**

Διαχείριση Δεδομένων
Διαμεσολάβησης



ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ ΔΙΑΜΕΣΟΛΑΒΗΣΗΣ

DATA ANALYSIS for MEDIATION

Ότι μπορεί να μετρηθεί δεν αξίζει πάντα και ότι αξίζει δεν μπορεί πάντα να μετρηθεί.



copyright@hitechPC

www.diamesolabisi.gr

Επί πλέον στο MADGROUP ανήκουν και τα πρωτότυπα λογισμικά

B) Το λογισμικό **DANSTEV**

Πρόγραμμα αξιολόγησης
προσωπικού μεγάλων
εταιρειών



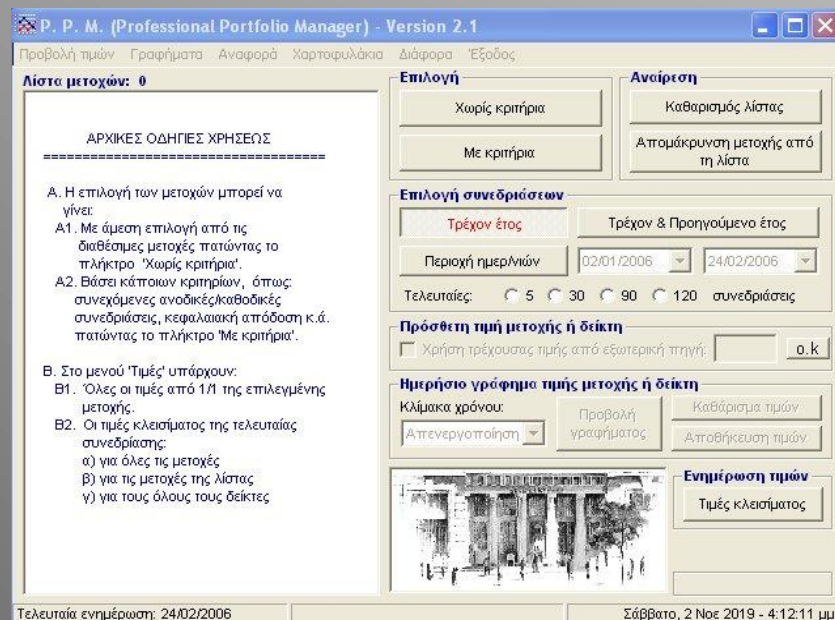
ANALYSE DE DONNEES D'EVOLUTION PROSOPHIQUE
DATA ANALYSIS for STAFF EVALUATION

Together
Evaluation
Archive
More
Whose
Opportunity
Responsibility
Kenly

copyright©hitechPC

Γ) Το Professional Portfolio Management **P.P.M**

Διαχειρίζεται μετοχές εισηγμένες στο Χ.Α.Α
και χαρτοφυλάκια μετοχών



P. P. M. (Professional Portfolio Manager) - Version 2.1

Προβολή τιμών Γραφήματα Αναφορά Χαρτοφυλάκια Διάφορα Έξοδος

Λίστα μετοχών: 0

ΑΡΧΙΚΕΣ ΟΔΗΓΙΕΣ ΧΡΗΣΕΩΣ

A. Η επιλογή των μετοχών μπορεί να γίνει:

A1. Με άμεση επιλογή από τις διαθέσιμες μετοχές πτατώντας το πλήκτρο 'Χωρίς κριτήρια'.

A2. Βάσει κάποιων κριτηρίων, όπως: συνεχόμενες ανοδικές/καθοδικές συνεδριάσεις, κεφαλαιακή απόδοση κ.ά. πτατώντας το πλήκτρο 'Με κριτήρια'.

B. Στο μενού 'Τιμές' υπάρχουν:

B1. Όλες οι τιμές από 1/1 της επιλεγμένης μετοχής.

B2. Οι τιμές κλεισίματος της τελευταίας συνεδρίασης:

α) για όλες τις μετοχές
β) για τις μετοχές της λίστας
γ) για τους όλους τους δείκτες

Επιλογή

Χωρίς κριτήρια
Με κριτήρια

Αναίρεση

Καθαρισμός λίστας
Απομάκρυνση μετοχής από τη λίστα

Επιλογή συνεδριάσεων

Τρέχον έτος Τρέχον & Προηγούμενο έτος

Περιοχή ημερ/μίων 02/01/2006 24/02/2006

Τελευταίες: 5 30 90 120 συνεδριάσεις

Πρόσθετη τιμή μετοχής ή δείκτη

Χρήση τρέχουσας τιμής από εξωτερική πηγή: o.k

Ημερήσιο γράφημα τιμής μετοχής ή δείκτη

Κλίμακα χρόνου: Προβολή γραφήματος Καθάρισμα τιμών

Απενεργοποίηση Αποθήκευση τιμών

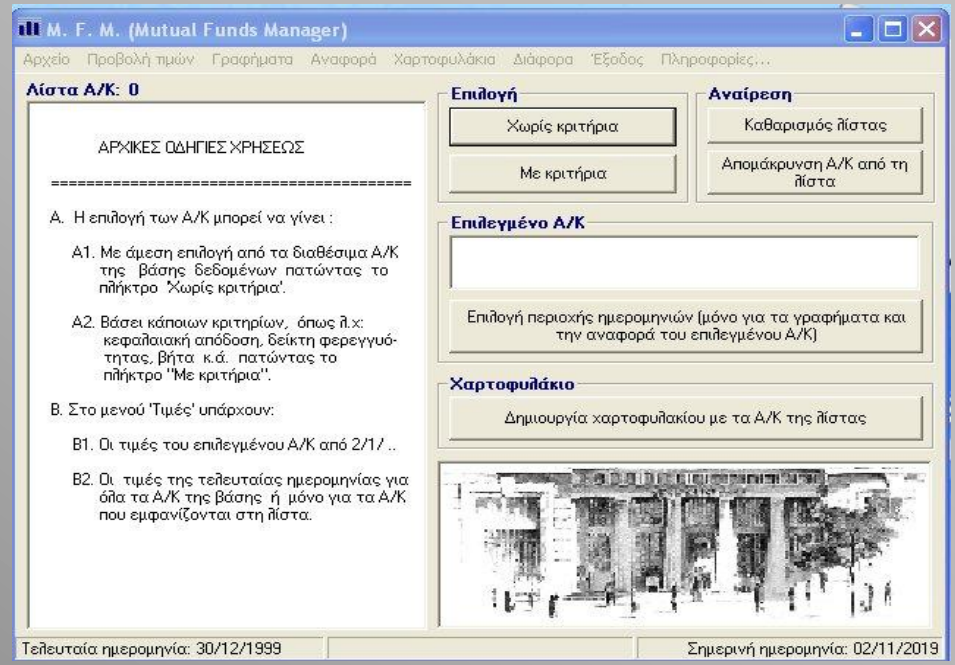
Ενημέρωση τιμών

Τιμές κλεισίματος

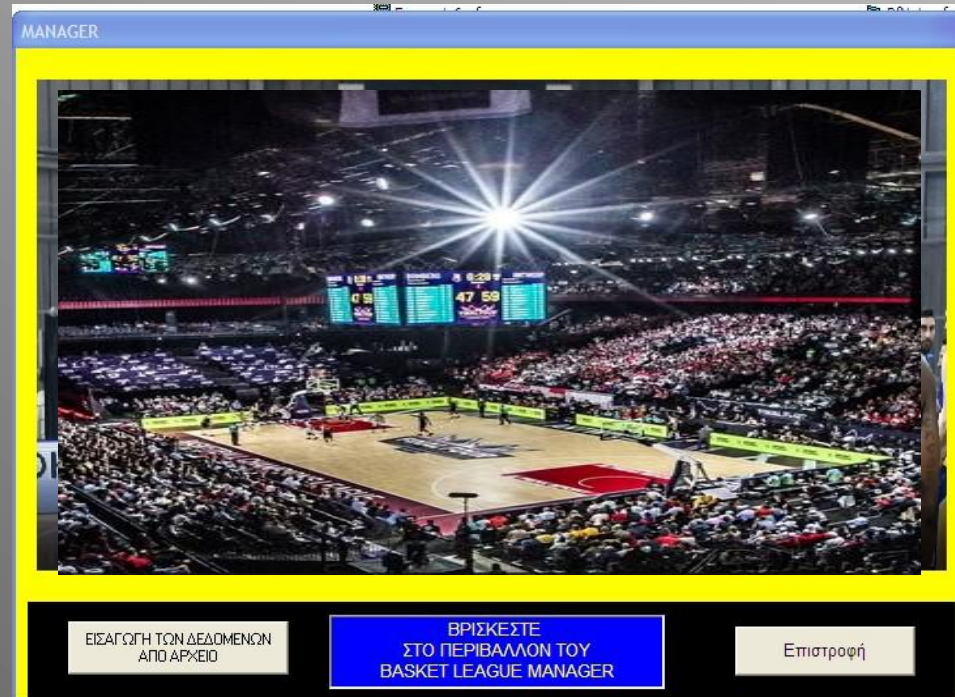
Τελευταία ενημέρωση: 24/02/2006 Σάββατο, 2 Νοε 2019 - 4:12:11 μμ

Τέλος στο MADGROUP ανήκουν και τα προγράμματα

Δ) Το Mutual Funds Management **M.F.M**
Διαχειρίζεται Αμοιβαία Κεφάλαια



Ε) Το Basket League Manager **BLM**
Αξιολογεί τις επιδόσεις παικτών ομάδων Μπάσκετ



Παραδείγματα εφαρμογών του λογισμικού MAD με τις πρωτότυπες μεθόδους KARAP και BENKAR του συγγραφέα

Θα παρουσιαστούν περιληπτικά πέντε εργασίες οι οποίες ολοκληρωμένες μπορεί κανείς να τις βρει στο βιβλίο του Δρος Δημητρίου Καραπιστόλη το οποίο περιλαμβάνει περισσότερες από 60 δημοσιευμένες εργασίες με τίτλο:

ΑΝΘΟΛΟΓΙΑ

ΕΦΑΡΜΟΓΩΝ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΜΕ ΠΟΛΥΠΑΡΑΓΟΝΤΙΚΕΣ ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΥΣ

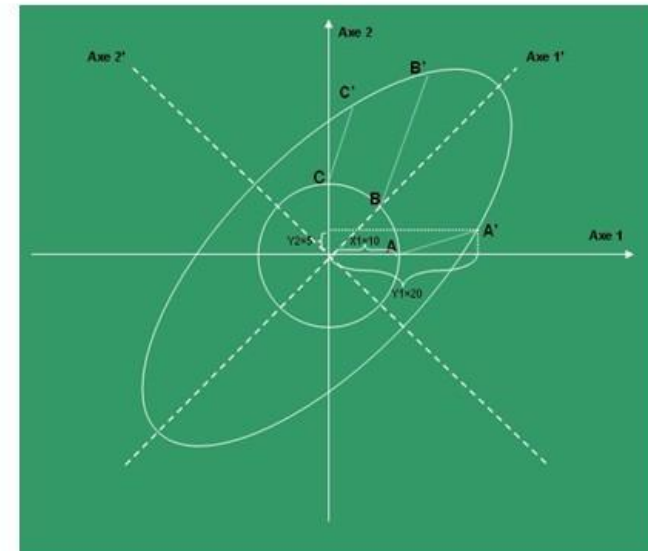
1. Η μέθοδος KARAP ως εργαλείο σχεδιασμού διαφημιστικής εκστρατείας ενός τουριστικού προορισμού
2. Η αξιολόγηση της εικόνας της Θεσσαλονίκης
3. Δημιουργία διαφημιστικού μηνύματος με τη μέθοδο KARAP
4. Χαρτογράφηση της ψυχολογίας των χρηστών INTERNET
- 5) Εντοπισμός ακραίων απαντήσεων ενός ερωτηματολογίου

Επικοινωνία: dimikarap@gmail.com

Δρ. ΔΗΜΗΤΡΙΟΣ ΝΙΚ. ΚΑΡΑΠΙΣΤΟΛΗΣ

ΑΝΘΟΛΟΓΙΑ

ΕΦΑΡΜΟΓΩΝ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΜΕ ΠΟΛΥΔΙΑΣΤΑΤΕΣ ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΥΣ



ΕΚΔΟΣΕΙΣ HITECH-PC

ΘΕΣΣΑΛΟΝΙΚΗ 2020

Α) Η ΜΕΘΟΔΟΣ ΚΑΡΑΡ ΩΣ ΕΡΓΑΛΕΙΟ ΣΧΕΔΙΑΣΜΟΥ ΔΙΑΦΗΜΙΣΤΙΚΗΣ ΕΚΣΤΡΑΤΕΙΑΣ ΕΝΟΣ ΤΟΥΡΙΣΤΙΚΟΥ ΠΡΟΟΡΙΣΜΟΥ

Στο ερώτημα του εντοπισμού ενός ερωτώμενου με ποιο κριτήριο συνδέεται κυρίως, θα γίνει λεπτομερής αναφορά, χρησιμοποιώντας ένα συγκεκριμένο ερωτηματολόγιο στο οποίο απάντησαν 231 Ρώσοι επισκέπτες από τους 1721 ξένους επισκέπτες της Θεσσαλονίκης από 51 χώρες της υφελίου, που ερωτήθηκαν την περίοδο 15-5-13 έως 15-9-13. Η επιλογή των συγκεκριμένων τουριστών έγινε επειδή το μέγεθός του δείγματος εκτός του ότι ήταν ικανοποιητικό, δηλαδή το 13,42% του συνόλου αλλά και λόγω ιδιαίτερων δεσμών που έχουν οι συγκεκριμένοι επισκέπτες με την πόλη. Τα δεδομένα περιέχονται στην έρευνα που διεξήχθη στα πλαίσια του προγράμματος ΑΡΧΙΜΗΔΗΣ ΙΙΙ με τίτλο «Τεχνολογίες Ανάλυσης Δεδομένων και Διαχείρισης Γνώσης στο σχεδιασμό τουριστικών προϊόντων» με **επιστημονικό υπεύθυνο τον καθηγητή Δρ. Δημήτριο Καραπιστόλη.**

Στα στοιχεία που θα επικεντρωθούμε είναι σχετικά με τις απαντήσεις Ρώσων επισκεπτών βάσει τριών ενότητων. Η πρώτη ενότητα αφορούσε στα **ΑΙΤΙΑ** που τους προκάλεσαν να επισκεφθούν την πόλη της Θεσσαλονίκης, ενώ η δεύτερη ενότητα αφορούσε στην κριτική τους για την **ΕΙΚΟΝΑ** που παρουσιάζει η πόλη. Η τρίτη ενότητα αφορούσε στο ποια θα είναι η **ΣΤΑΣΗ** που θα κρατήσουν στο μέλλον για μια εκ νέου επίσκεψη στη πόλη με τρεις διαβαθμίσεις (**Απίθανο-Πιθανό-Βέβαιο**).

Σχετικά με τα αίτια που προκάλεσαν την επίσκεψη χρησιμοποιήθηκαν τρεις ενότητες των πέντε κριτηρίων, όπου η ενότητα Γ1 η οποία αφορούσε τους λόγους επιλογής του προορισμού περιλάμβανε τα εξής πέντε κριτήρια

Γ11={Η φήμη του ως τουριστικού προορισμού}, Γ12={Οι φυσικές ομορφιές της περιοχής}, Γ13={Οι γνώμες φίλων και γνωστών}, Γ14={Το κλίμα} και Γ15={Ιστορία της περιοχής}.

Η Γ2 ενότητα αφορούσε στα στοιχεία που τους προσέλκυαν περισσότερο κατά την επίσκεψή τους και περιλάμβανε τα εξής πέντε κριτήρια.

Γ21={Μπάνια-Θαλάσσια σπόρ}, Γ22={Επισκέψεις μουσείων}, Γ23={Νυχτερινή διασκέδαση}, Γ24={Εκδρομές-Επαφές με τη φύση} και Γ25={Η τοπική κουζίνα}.

Η Γ3 ενότητα αφορούσε στους πιο σημαντικούς παράγοντες ενός προορισμού και περιλάμβανε τα εξής πέντε κριτήρια. Γ31={Οργάνωση-Υποδομές}, Γ32={Τοπικές μεταφορές}, Γ33={Ασφάλεια}, Γ34={Φυσικό περιβάλλον} και Γ35={Life Style}.

Σχετικά με την εικόνα της πόλης χρησιμοποιήθηκαν τα εξής εννέα κριτήρια. α) Δ1={Η καθαριότητα}, Δ2={Οι φυσικές ομορφιές}, Δ3={Οι τιμές προϊόντων και υπηρεσιών}, Δ4={Τα αξιοθέατα της πόλης της Θεσσαλονίκης}, Δ5={Η Ελληνική κουζίνα}, Δ6={Η νυχτερινή ζωή της πόλης}, Δ7={Το αρχιτεκτονικό της στυλ}, Δ8={Η ασφάλειά της} και Δ9={Η φιλικότητα των ντόπιων}

Πίνακας 1: Τμήμα του πίνακα των κωδικοποιημένων απαντήσεων

<u>ind</u>	<u>Γ1</u>	<u>Γ2</u>	<u>Γ3</u>	<u>Δ1</u>	<u>Δ2</u>	<u>Δ3</u>	<u>Δ4</u>	<u>Δ5</u>	<u>Δ6</u>	<u>Δ7</u>	<u>Δ8</u>	<u>Δ9</u>	<u>ΣΤ</u>
35	3	1	5	3	4	4	4	4	5	0	4	3	1
112	1	1	4	3	5	3	4	5	3	4	3	5	2
131	5	3	5	3	5	5	5	5	5	5	3	5	3
133	5	3	5	3	4	5	5	5	5	4	4	5	3
134	1	3	3	3	4	4	4	5	5	4	5	5	3
135	5	3	5	3	5	4	4	5	5	5	3	5	3

Με την εφαρμογή της μεθόδου KARAP έχουμε τον πίνακα των 231 Ρώσων που συνδέονται με τις 15 διαβαθμίσεις της 1^{ης} ενότητας των αιτίων της επίσκεψης.

Πίνακας 3: Ο πίνακας ταξινόμησης των ερωτώμενων

<u>IND</u>	<u>Γ11</u>	<u>Γ12</u>	<u>Γ13</u>	<u>Γ14</u>	<u>Γ15</u>	<u>Γ21</u>	<u>Γ22</u>	<u>Γ23</u>	<u>Γ24</u>	<u>Γ25</u>	<u>Γ31</u>	<u>Γ32</u>	<u>Γ33</u>	<u>Γ34</u>	<u>Γ35</u>
	23	5	17	22	8	10	25	11	33	13	14	10	19	4	17
1	112	417	320	327	220	425	209	131	207	313	188	169	477	307	35
2	365	420	371	359	339	618	214	133	228	743	340	419	592	478	389
3	366	459	372	370	426	983	215	134	229	845	424	456	674	847	543
4	376	612	373	451	852	993	256	135	238	859	672	640	750	1057	546
5	488	1020	1056	466	887	1003	353	460	293	888	848	1186	791		849
6	544		1444	510	964	1007	368	495	294	1059	963	1195	846		961
7	545		1508	585	1044	1046	374	513	295	1167	1010	1370	858		1006
8	635		1544	596	1247	1101	378	514	296	1389	1189	1406	1019		1023
9	673		1567	857		1118	418	1400	297	1390	1265	1474	1050		1115
10	676		1574	1008		1363	427	1509	298	1393	1330	1481	1081		1180
11	751		1618	1011			476	1668	299	1465	1625		1194		1181
12	790		1624	1047			749		300	1480	1660		1371		1409
13	850		1627	1048			752		301	1510	1667		1466		1476
14	908		1647	1062			851		302		1669		1467		1478
15	968		1654	1254			860		303				1485		1479
16	1083		1656	1256			962		304				1486		1511
17	1085		1659	1261			965		305				1614		1626
18	1193			1264			1009		306				1615		
19	1213			1564			1021		308				1655		
20	1280			1613			1069		309						
21	1374			1657			1188		310						
22	1383			1671			1249		311						
23	1408						1279		312						
24							1619		314						
25							1653		315						
26									319						
27									367						
28									909						
29									1196						
30									1278						
31									1443						
32									1475						
33									1484						

Κατανομή

15 διαβαθμίσεων της ενότητας των Αιτίων ως προς τις 3 Στάσεις

IND	ΣΤ1	ΣΤ2	ΣΤ3
Γ11	2	14	7
Γ12	1	2	2
Γ13	0	6	11
Γ14	4	11	7
Γ15	1	4	3
Γ21	0	8	2
Γ22	3	14	8
Γ23	1	6	4
Γ24	0	18	15
Γ25	1	7	5
Γ31	2	11	1
Γ32	0	3	7
Γ33	0	11	8
Γ34	0	3	1
Γ35	2	10	5
	17	128	86

Κατανομή

45 διαβαθμίσεων των διαβαθμίσεων της ΕΙΚΟΝΑΣ ως προς τις 3 Στάσεις

IND	ΣΤ1	ΣΤ2	ΣΤ3
Δ14	0	1	4
Δ24	2	21	3
Δ25	1	7	16
Δ31	0	1	0
Δ33	5	9	4
Δ34	0	6	2
Δ44	0	6	0
Δ45	0	17	22
Δ54	1	3	0
Δ55	1	15	10
Δ61	2	0	0
Δ63	0	1	0
Δ64	0	1	2
Δ72	0	1	0
Δ74	3	8	4
Δ75	0	2	5
Δ83	0	4	0
Δ84	1	4	1
Δ85	0	1	0
Δ94	0	0	1
Δ95	1	20	12
	17	128	86

Πλήθος ατόμων ως προς τις Διαβαθμίσεις των ενότητων ΑΙΤΙΑ και ΕΙΚΟΝΑ τις τρεις ΣΤΑΣΕΙΣ

ΣΤΑΣΗ 1				ΣΤΑΣΗ 2				ΣΤΑΣΗ 3			
Γ12	1	Δ33	5	Γ11	14	Δ24	21	Γ13	11	Δ14	4
		Δ61	2	Γ14	11	Δ31	1	Γ24	15	Δ25	16
				Γ15	4	Δ34	6	Γ32	7	Δ45	22
				Γ21	8	Δ44	6	Γ33	8	Δ64	2
				Γ22	14	Δ54	3			Δ75	5
				Γ23	6	Δ55	15			Δ94	1
				Γ25	7	Δ63	1				
				Γ31	11	Δ72	1				
				Γ34	3	Δ74	8				
				Γ35	10	Δ83	4				
						Δ84	4				
						Δ85	1				
						Δ95	20				
	1		7		88		91		41		50

Πίνακας 13: Κατανομή ερωτώμενων (και απαντήσεων) για κάθε περίπτωση

Περιπτώσεις	ΣΤΑΣΗ 1	ΣΤΑΣΗ 2	ΣΤΑΣΗ 3	ΣΥΝΟΛΟ
Μόνο με κριτήρια των αιτίων	1	25	14	40
Μόνο με κριτήρια της εικόνα	7	28	23	58
Και με τα δύο κριτήρια	0	63 (2x63=126)	27 (2x27=54)	90 (2x90=180)
Με τα άλλα κριτήρια κάθε στάσης	9	12	22	43
ΣΥΝΟΛΟ (Επισκεπτών)	17	128	86	231
ΣΥΝΟΛΟ (Απαντήσεων)	(8)	(179)	(91)	(278)

Αρχικά από τον πίνακα 13 εντοπίζεται η πιθανότητα (άρα και το ποσοστό) κάθε στάσης. Έτσι για την Στάση 1, όσοι θεωρούν Αλίθανο να επισκεφτούν την πόλη, το ποσοστό είναι 7,36%, για όσους είναι Πιθανό (Στάση 2) το ποσοστό ανέρχεται στο 55,41%, ενώ όσοι είναι βέβαιοι ότι θα επαναλάβουν μια νέα επίσκεψη (Στάση 3) ανέρχονται στο 37,23%. Ακολουθεί ο πίνακας 14, ο οποίος παρουσιάζει την πιθανότητα κάθε στάσης με βάση τις απαντήσεις σύμφωνα με τα δεδομένα του πίνακα 13, λαμβάνοντας υπόψη τα άτομα που εντοπίζονται για κάθε περίπτωση

Πίνακας 14: Πίνακας πιθανοτήτων για κάθε περίπτωση και κάθε στάση

Περιπτώσεις	ΣΤΑΣΗ 1	ΣΤΑΣΗ 2	ΣΤΑΣΗ 3
Μόνο από τις αιτίες	0,0588	0,1953	0,1628
Μόνο από την εικόνα	0,4118	0,2188	0,2675
Και από τα δύο κριτήρια	0	0,4921	0,3139
Με τα άλλα κριτήρια κάθε στάσης	0,5294	0,0938	0,2558
Σύνολο	1	1	1

Εφαρμόζοντας στα δεδομένα του πίνακα 13 το τεστ ανεξαρτησίας σε επίπεδο σημαντικότητας $\alpha=5\%$, έδειξε το εξής αποτέλεσμα.

βαθμοί ελευθερίας $v=6$, $\chi^2(v,0.05)=12,591$ και $\chi^2=34,289$. Επειδή $\chi^2 > \chi^2(v,0.05)$ συνεπάγεται ότι η στάση ενός επισκέπτη **εξαρτάται από τα κριτήρια που διαμορφώνουν την επιλογή του να επισκεφθεί τη Θεσσαλονίκη.**

ΣΥΜΠΕΡΑΣΜΑΤΑ

Από το γενικό ποσοστό 7,36% των Ρώσων που δήλωσαν αρνητική στάση, ξεχωρίζει ένα πολύ μικρό ποσοστό (5,88%) το οποίο δεν θα επαναλάβει την επίσκεψη, επειδή δεν άρεσε τις φυσικές ομορφιές της περιοχής, ενώ ένα ποσοστό 41,18% δεν έχει καθόλου καλή εικόνα της Θεσσαλονίκης για τις τιμές των προϊόντων και τη νυκτερινή ζωή της πόλης.

Από το γενικό ποσοστό 55,41% των Ρώσων που δήλωσαν ότι είναι πιθανή μια επανάληψη της επίσκεψης, ξεχωρίζει ένα σημαντικό ποσοστό (49,21%) το οποίο ως αίτιο προσέλκυσης υπήρξε η φήμη του προορισμού, το κλίμα, οι επισκέψεις μουσείων, η οργάνωση-υποδομές της περιοχής και το Life Style, ενώ ενθουσιάστηκαν κυρίως με τις φυσικές ομορφιές της πόλης, με την Ελληνική κουζίνα και την φιλικότητα των ντόπιων.

Από το γενικό ποσοστό 37,23% των Ρώσων που δήλωσαν ότι είναι βέβαιη μια επανάληψη της επίσκεψης ξεχωρίζει ένα ποσοστό (31,39%) το οποίο ως αίτιο προσέλκυσης υπήρξαν οι γνώμες φίλων και γνωστών, καθώς και η δυνατότητα που παρέχει η περιοχή για εκδρομές, ενώ ενθουσιάστηκαν με τις φυσικές ομορφιές και με τα αξιοθέατα της πόλης. Επίσης ένα ικανοποιητικό ποσοστό 26,75% στηρίζει την συγκεκριμένη απόφαση επειδή έμειναν ικανοποιημένοι από την εικόνα της πόλης.

Λόγω του υψηλού ποσοστού (92,64%) των Ρώσων που έχουν θετική και μάλλον θετική στάση για μια μελλοντική επίσκεψη στη Θεσσαλονίκη, μια σωστή διαφημιστική εκστρατεία για τους ομοεθνείς τους, πρέπει να εστιαστεί κυρίως στα παρακάτω θετικά στοιχεία που προσδιόρισαν οι Ρώσοι επισκέπτες της Θεσσαλονίκης.

Έτσι

A) Όσον αφορά στις αιτίες που θα προκαλούσαν μια επίσκεψη στη πόλη, πρέπει να δοθεί έμφαση στη φήμη που έχει η πόλη ως τουριστικός προορισμός, στο κλίμα της περιοχής, στις επισκέψεις μουσείων, στο Life Style της πόλης και στη δυνατότητα για εκδρομές στη φύση.

B) Σε ότι έχει σχέση με την εικόνα της πόλης πρέπει να τονιστούν τα αξιοθέατα της πόλης, η Ελληνική κουζίνα και η φιλικότητα των ντόπιων.

Η εφαρμογή της μεθόδου KARAP στη διαμόρφωση του σχεδιασμού της διαφημιστικής εκστρατείας ενός τουριστικού προορισμού μετά από μία έρευνα αγοράς με κριτήρια που προσδιορίζουν τον λόγο της επίσκεψης των ερωτώμενων, αλλά και με κριτήρια που καθορίζουν την εικόνα της περιοχής, προσδιορίζει με ακρίβεια

A) την μελλοντική στάση της επίσκεψης από τους ίδιους ερωτηθέντες

B) Ποια κριτήρια χαρακτηρίζουν την περιοχή

B) ΑΞΙΟΛΟΓΗΣΗ ΤΗΣ ΕΙΚΟΝΑΣ ΤΗΣ ΘΕΣΣΑΛΟΝΙΚΗΣ

Στην περίπτωση αυτή ως στοιχεία θα χρησιμοποιηθούν οι απαντήσεις των 1721 επισκεπτών από τις 51 χώρες του πλανήτη που επισκέφθηκαν την Θεσσαλονίκη, που αφορούν την ΕΙΚΟΝΑ της Πόλης με τις 9 διαβαθμίσεις χωρισμένες σε δύο γκρουπ απαντήσεων προς αξιολόγηση των δύο ταξινομήσεων που εφαρμόζονται στους 1721 επισκέπτες. Το 1^ο γκρουπ με τις διαβαθμίσεις Δ1 Έως Δ3 (όπου Δ1=την καθαριότητα της πόλης, Δ2=τις φυσικές ομορφιές και Δ3=τις τιμές των προϊόντων και υπηρεσιών) και το 2^ο γκρουπ με τις διαβαθμίσεις Δ4-Δ9 (όπου Δ4= τα αξιοθέατα της πόλης της Θεσσαλονίκης Δ5= η Ελληνική κουζίνα Δ6= η νυχτερινή ζωή της πόλης Δ7= το αρχιτεκτονικό της στυλ της Πόλης Δ8= η ασφάλειά της Πόλης και Δ9= η φιλικότητα των ντόπιων. .

Στον πίνακα 7 έχουμε τις τρεις κατανομές σε πέντε κλάσεις που προήλθαν μετά από ταξινόμηση των 1721 ατόμων

α) με την Ανιούσα Ιεραρχική Ταξινόμηση

β) με την μέθοδο KARAP (DIS) στο χώρο R⁵ που δημιούργησαν οι πέντε παραγοντικοί άξονες μετά την εφαρμογή της Παραγοντικής Ανάλυσης των Αντιστοιχιών (-AFC-) στον πίνακα δεδομένων T(1726,6)

γ) την κατανομή των μέγιστων πιθανοτήτων να ανήκουν τα 1721 άτομα σε ΜΙΑ από τις πέντε διαφορετικές κλάσεις της Ανιούσας Ιεραρχικής Ταξινόμησης (-CAH-).

Όσον αφορά τα κριτήρια Δ1 έως Δ3

Πίνακας 6: Τιμές των τριών μεταβλητών και οι κλάσεις στις οποίες ανήκουν οι ερωτώμενοι μετά την ταξινόμηση με την διαδικασία FACOR

IND	Δ1	Δ2	Δ3	Κλάση FACOR
I1	3	5	4	5
I2	2	5	3	5
I3	1	3	4	5
.				
1690	2	3	5	5
.				
1719	1	2	3	5
1720	1	5	2	3
1721	2	5	3	5

Από το λογισμικό MAD δίδονται τα ακόλουθα αποτελέσματα:

Τακτοποιήθηκαν στις ΙΔΙΕΣ κλάσεις :1638 άτομα

Τακτοποιήθηκαν σε ΔΙΑΦΟΡΕΤΙΚΕΣ κλάσεις : 83 άτομα

Ποσοστό καλής προσαρμογής : 95,18%

Πίνακας 7: Οι τρεις κατανομές απολύτων συχνοτήτων του 7^{ου} βήματος

CAH	n _i	f _i	DIS	n _i	f _i	Κατανομή πιθανοτήτων	n _i
K1	119	0,0691	K1	135	0,0784	T1: 0,3700 - 0,4970	68
K2	294	0,1708	K2	348	0,2022	T2: 0,4970 - 0,6241	201
K3	221	0,1284	K3	219	0,1272	T3: 0,6241 - 0,7512	161
K4	691	0,4015	K4	681	0,3957	T4: 0,7512 - 0,8784	254
K5	396	0,23	K5	338	0,1963	T5: 0,8784 - 1,0000	954
	1721	1		1721	1		1638

Προχωρώντας στο βήμα 8 η αξιολόγηση της συγκεκριμένης ταξινόμησης των 1721 επισκεπτών με την μέθοδο BENKAR με βάση τα στοιχεία του πίνακα 7 προκύπτει η εξής αξιολόγηση: από τα 1721 άτομα τα 1208=(954+254) , δηλαδή το 70,18% του συνόλου των ερωτηθέντων έχει πιθανότητα από 0,7512 και άνω να ανήκει σε ΜΙΑ από τις πέντε κλάσεις της ταξινόμησης με την μέθοδο FACOR.

Επειδή το ποσοστό 70,18% του συνόλου των ερωτηθέντων είναι ικανοποιητικό, ενώ και το ποσοστό 95,18% της προσαρμογής των «αντικειμένων» στις πέντε κλάσεις με βάση την ταξινόμηση FACOR είναι υψηλό, τα δεδομένα του πίνακα T(1721,3) εντός των κλάσεων αυτών μπορεί να θεωρηθούν ότι έχουν την απαραίτητη επιθυμητή ομοιογένεια, σε σχέση με τις τιμές των τριών μεταβλητών.

Παρατήρηση: Η ερμηνεία της κλάσης T5 είναι η εξής: 954 άτομα από τα 1638 που τακτοποιήθηκαν στις ίδιες κλάσεις, δηλαδή ποσοστό 58,24% έχει πιθανότητα από 0,8774 και άνω να ανήκει σε ΜΙΑ από τις πέντε κλάσεις της ταξινόμησης CAH.

Από το λογισμικό MAD δίδονται τα ακόλουθα αποτελέσματα:

Τακτοποιήθηκαν στις ΔΙΕΣ κλάσεις :1416 άτομα

Τακτοποιήθηκαν σε ΔΙΑΦΟΡΕΤΙΚΕΣ κλάσεις : 305 άτομα

Ποσοστό καλής προσαρμογής : **82,28%**

Πίνακας 5: Οι τρεις κατανομές του 7^{ου} βήματος

CAH	n _i	f _i	DIS	n _i	f _i	Probability distribution	ni
K1	99	0,0575	K1	136	0,079	T1: 0,2305 - 0,3855	48
K2	181	0,1051	K2	225	0,1307	T2: 0,3855 - 0,5408	230
K3	129	0,0749	K3	142	0,0825	T3: 0,5408 - 0,6962	315
K4	848	0,4927	K4	764	0,4439	T4: 0,6962 - 0,8515	335
K5	464	0,2696	K5	454	0,2638	T5: 0,8515 - 1,0000	488
	1721	1		1721	1		1416

Παρατήρηση: Η ερμηνεία της κλάσης T5 είναι η εξής: **488** άτομα από τα **1416** που τακτοποιήθηκαν στις ίδιες κλάσεις, δηλαδή ποσοστό **34,5%** έχει πιθανότητα από **0,8515** και άνω να ανήκει σε ΜΙΑ από τις πέντε κλάσεις της ταξινόμησης DIS.

Προχωρώντας στο βήμα 8 η αξιολόγηση της συγκεκριμένης ταξινόμησης των 1721 επισκεπτών με την μέθοδο BENKAR, χρησιμοποιώντας τις δύο τελευταίες τάξεις της κατανομής πιθανοτήτων είναι η εξής: 823 (=488+335) άτομα από τα 1721, δηλαδή το 47,82% του συνόλου των ερωτηθέντων έχει πιθανότητα από 0,6962 και άνω να ανήκει σε ΜΙΑ από τις πέντε κλάσεις της ταξινόμησης.

Επειδή το ποσοστό 47,82% του συνόλου των ερωτηθέντων δεν είναι ικανοποιητικό, παρά το ότι το ποσοστό 82,28% της προσαρμογής των «αντικειμένων» στις πέντε κλάσεις των δύο ταξινομήσεων είναι αρκετά υψηλό, τα δεδομένα του πίνακα T(1721,6) εντός των κλάσεων φαίνεται ότι δεν έχουν την απαραίτητη επιθυμητή ομοιογένεια, σε σχέση με τις τιμές των έξι μεταβλητών.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Συγκρίνοντας τις δύο αξιολογήσεις των ταξινομήσεων αφενός με τα κριτήρια Δ1-Δ3, αφετέρου με τα κριτήρια Δ4-Δ9, αναδεικνύεται ότι οι 1721 επισκέπτες της Θεσσαλονίκης είχαν **ομοιόμορφη εικόνα** της πόλης ως προς τα τρία πρώτα κριτήρια που κλήθηκαν να αξιολογήσουν, ενώ για τα υπόλοιπα έξι κριτήρια δεν παρουσίαζαν ανάλογη **ομοιογένεια** απαντήσεων. Αυτό μπορεί να οφείλεται επειδή οι επισκέπτες προερχόταν από 51 διαφορετικές χώρες του πλανήτη, οπότε κρίνεται φυσιολογικό στα κριτήρια Δ1 έως Δ3 να έχουν περίπου την ίδια αντίληψη, ενώ σε κάποια από τα κριτήρια Δ4 έως Δ9 να παρουσιάζουν τελείως διαφορετικές απόψεις λόγω διαφορετικών παραδόσεων που ισχύουν στον τόπο του κάθε επισκέπτη.

ΕΚΠΑΙΔΕΥΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΗΝ ΜΗΧΑΝΗ ΔΙΑΝΥΣΜΑΤΩΝ ΥΠΟΣΤΗΡΙΞΗΣ –SVM-

Εκπαιδεύοντας με 20 επαναλήψεις τα δεδομένα του πίνακα ταξινόμησης με τις τιμές των 6 διαβαθμίσεων Δ4-Δ9 και κάνοντας χρήση της Μηχανής Διανυσμάτων Υποστήριξης (-SVM-), διατηρώντας σε κάθε επανάληψη ένα τυχαίο δείγμα 20% των 1721 τιμών, μια φορά με τις κατατάξεις των «αντικειμένων» με την μέθοδο FACOR και την άλλη με την μέθοδο BENKAR έχουμε τα ακόλουθα αποτελέσματα:

Πίνακας 8: Ποσοστά εκμάθησης των 20 επαναλήψεων μετά την εκπαίδευση των ταξινομήσεων των δεδομένων βάσει της μεθόδου FACOR και της μεθόδου BENKAR

FACOR (20%)		BENKAR (20%)	
0,6950		0,8363	
0,6862		0,9620	
0,7595		0,7105	
0,7830		0,9415	
0,7009		0,9444	
0,7185		0,9094	
0,7859		0,7515	
0,8211		0,9532	
0,7889	76,54%	0,6462	87,88%
0,8710		0,9035	
0,8328		0,7076	
0,8182		0,9064	
0,6862		0,9474	
0,7830		0,9474	
0,7067		0,8918	
0,8065		0,9737	
0,8035		0,9386	
0,7155		0,7982	
0,7595		0,9561	
0,7859		0,9503	

Σημείωση: Από τα 350 άτομα του δείγματος μόνο το 69,5% τοποθετήθηκε στις κλάσεις του υποδείχθηκε από την ταξινόμηση FACOR

Σημείωση: Από τα 350 άτομα του δείγματος το 83,63% τοποθετήθηκε στις κλάσεις του υποδείχθηκε από την ταξινόμηση BENKAR

Από τον πίνακα 8 διαπιστώνεται ότι η μέθοδος BENKAR υπερέχει στην σωστή ταξινόμηση των αντικειμένων. Αυτό προκύπτει επειδή το ποσοστό εκτίμησης της εκμάθησης των δεδομένων του πίνακα 1, που αφορά την ταξινόμηση με την μέθοδο BENKAR (87,88%) είναι υψηλότερο από εκείνο που προκύπτει με την Ανιούσα Ιεραρχική Ταξινόμηση με την μέθοδο FACOR (76,54%). Επί πλέον με την μέθοδο BENKAR στις 20 επαναλήψεις εκμάθησης των δεδομένων τα ποσοστά εκτίμησης πάνω από την μέση τιμή είναι κατά πολύ υψηλότερα (13 στις 20 επαναλήψεις πάνω από 90% με μέγιστη τιμή 97,37%) από τα αντίστοιχα ποσοστά εκτίμησης με την μέθοδο FACOR που η μέγιστη τιμή ανέρχεται μόλις στο 87,10%.

Με δεδομένο ότι όπως προαναφέρθηκε η SVM δεν επιστρέφει πιθανότητες, ενώ η μέθοδος BENKAR υπολογίζει την πιθανότητα κάθε στατιστικής μονάδας να ανήκει σε μία ορισμένη κλάση, ως εκ τούτου η κατανομή των μέγιστων πιθανοτήτων που προκύπτει από την προτεινόμενη μέθοδο, μπορεί να θεωρηθεί ότι αξιολογεί αντικειμενικά την Ανιούσα Ιεραρχική Ταξινόμηση που απορρέει με την μέθοδο FACOR.

Γ) ΔΗΜΙΟΥΡΓΙΑ ΔΙΑΦΗΜΙΣΤΙΚΟΥ ΜΗΝΥΜΑΤΟΣ ΜΕ ΤΗΝ ΜΕΘΟΔΟ KARAP

◇ Στόχοι της έρευνας

Το θέμα της εφαρμογής αφορά στην δημιουργία διαφημιστικού μηνύματος για ένα συγκεκριμένο προϊόν, τα «τσιπς». Φυσικά η μεθοδολογία μπορεί να χρησιμοποιηθεί για οποιοδήποτε προϊόν, αρκεί να ορισθούν τα κατάλληλα κριτήρια (ποιοτικές μεταβλητές) που θα περιγράφουν το δυνατόν καλύτερα, το πλαίσιο μέσα στο οποίο πρέπει να κινηθεί η διαφημιστική εκστρατεία.

Στην συγκεκριμένη περίπτωση στόχοι της έρευνας είναι να γίνουν κατανοητές οι εξής καταστάσεις:

- ◇ Να διαπιστωθούν οι προτιμήσεις των καταναλωτών σχετικά με τα σημεία πώλησης των τσιπς.
- ◇ Να ερευνηθούν οι γεύσεις των τσιπς που σημειώνουν την μεγαλύτερη προτίμηση των καταναλωτών.
- ◇ Να διερευνηθούν οι ώρες που συνήθως καταναλώνουν τα τσιπς.
- ◇ Να διαπιστωθεί ο βαθμός ικανοποίησης για τις γεύσεις που συνήθως καταναλώνουν οι αγοραστές των τσιπς.
- ◇ Να διερευνηθούν οι συγκεκριμένοι λόγοι που επηρεάζουν την αγορά του προϊόντος.

Παρουσίαση του ερωτηματολογίου

Ερώτηση 1^η: Ποιο σημείο πώλησης προτιμάτε συνήθως για την αγορά του προϊόντος;

Περίπτερα Ψιλικά Super Market

Ερώτηση 2^η: Επιλέξτε ποιες από τις παρακάτω γεύσεις προτιμάτε;

Φυσική Ρίγανη Πάπρικα Τάκος

Ερώτηση 3^η: Ποιες ώρες συνήθως καταναλώνετε τα τσιπς;

Πρωί Απόγευμα Βράδυ

Ερώτηση 4^η : Εκφράστε το βαθμό ικανοποίησής σας για την ποιότητα του προϊόντος που συνήθως καταναλώνετε;

Μέτριος Μεσαίος Μεγάλος

Ερώτηση 5^η : Ποιος από τους παρακάτω λόγους επηρεάζει συνήθως την επιλογή σας για το συγκεκριμένο προϊόν.

Τιμή Γεύση Διαφήμιση

Οι ερωτήσεις που παρατίθενται στο ερωτηματολόγιο εκφράζονται με ποιοτικές μεταβλητές και χωρίζονται σε συγκεκριμένες διαβαθμίσεις και απαντήθηκε από 50 νέους έως 35 ετών.

Ταξινόμηση των ερωτώμενων με βάση της μέθοδο KARAP

ΣΠ1	ΣΠ2	ΣΠ3	ΓΕ1	ΓΕ2	ΓΕ3	ΓΕ4	ΩΡ1	ΩΡ2	ΩΡ3	ΙΚ1	ΙΚ2	ΙΚ3	ΕΠ1	ΕΠ2	ΕΠ3
4	7	1	0	3	0	0	4	5	1	2	4	5	0	3	11
13	19	142		16			110	11	145	146	12	114		15	14
111	117			137			120	17		149	112	115		18	116
123	124			143			147	113			118	120		131	121
135	127						148	119			144	132			122
	128							125				136			126
	133														129
	138														134
															139
															140
															141
															150

Με βάση την ταξινόμηση και το πλήθος των ερωτώμενων σε κάθε διαβάθμιση οι σημαντικότερες διαβαθμίσεις κάθε μεταβλητής είναι :

Όσον αφορά στην μεταβλητή «Σημείο πώλησης (ΣΠ)»: Είναι η διαβάθμιση «Ψιλικά -ΣΠ2»

Όσον αφορά στην μεταβλητή «Γεύσεις (ΓΕ)»: Είναι η διαβάθμιση «Ρίγανη-ΓΕ2»

Όσον αφορά στην μεταβλητή «Ωρες κατανάλωσης (ΩΡ)»: Είναι η διαβάθμιση «Απόγευμα-ΩΡ2»

Όσον αφορά στην μεταβλητή «Βαθμός ικανοποίησης (ΙΚ)»: Είναι η διαβάθμιση «Μεγάλη ικανοποίηση-ΙΚ3»

Όσον αφορά στην μεταβλητή «Λόγοι επιλογής (ΕΠ)»: Είναι η διαβάθμιση «Διαφήμιση-ΕΠ3»

Με βάση τα συμπεράσματα που προέκυψαν από την ταξινόμηση των ερωτώμενων χρησιμοποιώντας την μέθοδο KARAP, τα **χαρακτηριστικά στοιχεία** στα οποία θα πρέπει να βασιστεί ο **σχεδιασμός του διαφημιστικού μηνύματος** είναι τα εξής:

- Είδος γεύσης → Ρίγανη
- Σημείο αγοράς → Ψιλικά
- Ζώνη προβολής → Απόγευμα
- Χρήση προϊόντος → Μεγάλη ικανοποίηση
- Επηρεασμός γνώμης → Διαφήμιση

ΠΡΟΤΕΙΝΟΜΕΝΟ ΔΙΑΦΗΜΙΣΤΙΚΟ ΜΗΝΥΜΑ

Το διαφημιστικό μήνυμα **ΒΑΣΙΖΟΜΕΝΟ** στα χαρακτηριστικά στοιχεία θα περιλαμβάνει τις εξής εικόνες:

Εικόνα 1η :Είναι απόγευμα, δύο νεαρά αδέρφια, ένα αγόρι και ένα κορίτσι, κάθονται στο σαλόνι του σπιτιού τους μπροστά στην Τ.Υ συζητώντας για το πάρτυ των γενεθλίων του αδελφού που πρόκειται να πραγματοποιηθεί το ίδιο βράδυ. Κεντρικό τους πρόβλημα είναι το τι θα προσφέρουν στους νεαρούς φίλους τους. Την προσοχή τους όμως την αποσπά μια όμορφη κοπέλα που διαφημίζει τσιπς με γεύση Ρίγανη της εταιρείας «X».

Εικόνα 2η :Με το τέλος της διαφήμισης τα δύο αδέρφια τρέχουν στο μαγαζί με Ψιλικά της γειτονιάς και αγοράζουν όλα τα σακουλάκια τσιπς με γεύση ρίγανη της εταιρείας "X".

Εικόνα 3η : Η ώρα του πάρτι έχει φθάσει. Όλοι διασκεδάζουν και τα πατατάκια με γεύση ρίγανη της εταιρείας «X» είναι παντού. Ξαφνικά το κουδούνι της εξώπορτας χτυπά. Όλοι αναρωτιούνται ποιος μπορεί να είναι αφού όλοι οι φίλοι είναι παρόντες.

Εικόνα 4η : Η πόρτα ανοίγει. Τα δύο αδέρφια τα χάνουν. Η πρωταγωνίστρια της διαφήμισης των τσιπς που πριν λίγο έβλεπαν στην Τ.Υ βρίσκεται στο κατώφλι της πόρτας τους. Εκείνη πλησιάζει προς το μέρος του αγοριού. Σκύβει στο τραπέζι με τα πατατάκια, παίρνει ένα από αυτά και το βάζει στο στόμα του.

Εικόνα 5η : Η κοπέλα ρίχνει την ατάκα: «Η γεύση της ρίγανης προσφέρει **ΠΟΛΛΕΣ ΣΥΓΚΙΝΗΣΕΙΣ**».

Συμπέρασμα

Η χρήση της μεθόδου **KARAP**, όπως αυτή υλοποιείται μέσα από το λογισμικό **MAD**, στην δημιουργία ενός διαφημιστικού μηνύματος προσφέρει με σαφήνεια την στρατηγική την οποία οφείλουν να ακολουθήσουν οι εμπλεκόμενοι στο τμήμα ενός δημιουργικού, αν βέβαια επιθυμούν να λάβουν υπόψη τους την άποψη του καταναλωτικού κοινού του προϊόντος που πρόκειται να διαφημίσουν.

Δ) ΧΑΡΤΟΓΡΑΦΗΣΗ ΤΗΣ ΨΥΧΟΛΟΓΙΑΣ ΤΩΝ ΧΡΗΣΤΩΝ ΤΟΥ INTERNET

Η χαρτογράφηση της ψυχολογίας μιας κοινωνικής ομάδας μπορεί να πραγματοποιηθεί κυρίως με την χρήση της μεθοδολογίας της **Σημειομετρίας** και της μεθόδου **KARAP**.

ΤΟ ΣΗΜΕΙΟΜΕΤΡΙΚΟ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

Το σημειομετρικό ερωτηματολόγιο, δεν αποσκοπεί να αποκομίσει **γνώμες** από ερωτήματα που παρουσιάζονται με μορφή προτάσεων, αλλά να **αποδώσει τιμές σε ορισμένο πλήθος λέξεων** ανάλογα με το ευχάριστο ή το δυσάρεστο συναίσθημα που προκαλεί η επίκλησή τους, ώστε να αναδυθεί η βασική **σημειομετρική δομή** του κοινωνικού συνόλου στο οποίο αναφέρεται.

Πίνακας 1: Η μορφή του σημειομετρικού ερωτηματολογίου

Ευρισκόμενοι στο Διαδικτυακό περιβάλλον βάλτε μέσα σε ένα κύκλο τον αντίστοιχο βαθμό για το συναίσθημα που σας προκαλεί κάθε μία από τις παρακάτω λέξεις. Παραδείγματος χάριν

◇ Για το πολύ δυσάρεστο συναίσθημα κυκλώστε το -3

◇ Για το πολύ ευχάριστο συναίσθημα κυκλώστε το 3

◇ Οι βαθμοί από -2 έως +2 κλιμακώνουν αυξητικά την ένταση του συναισθήματος που σας προκαλεί η λέξη

α/α	ΛΕΞΕΙΣ	ΒΑΘΜΟΙ						
Λ1	Άγχος	-3	-2	-1	0	1	2	3
Λ2	Αναγνώριση	-3	-2	-1	0	1	2	3
Λ3	Ανευθυνότητα	-3	-2	-1	0	1	2	3
Λ4	Ανοχή	-3	-2	-1	0	1	2	3
Λ5	Γνώση	-3	-2	-1	0	1	2	3
Λ6	Διάθεση	-3	-2	-1	0	1	2	3
Λ7	Διασκέδαση	-3	-2	-1	0	1	2	3
Λ8	Εγκατάλειψη	-3	-2	-1	0	1	2	3
Λ9	Εγκράτεια	-3	-2	-1	0	1	2	3
Λ10	Εικονική πραγματικότητα	-3	-2	-1	0	1	2	3
Λ11	Εκκλησία	-3	-2	-1	0	1	2	3
Λ12	Ελευθερία	-3	-2	-1	0	1	2	3
Λ13	Εξάρτηση	-3	-2	-1	0	1	2	3
Λ14	Επικοινωνία	-3	-2	-1	0	1	2	3
Λ15	Κατάθλιψη	-3	-2	-1	0	1	2	3
Λ16	Μανία	-3	-2	-1	0	1	2	3
Λ17	Μελαγχολία	-3	-2	-1	0	1	2	3
Λ18	Μοναξιά	-3	-2	-1	0	1	2	3
Λ19	Ναρκωτικά	-3	-2	-1	0	1	2	3
Λ20	Οικογένεια	-3	-2	-1	0	1	2	3
Λ21	Παρέα	-3	-2	-1	0	1	2	3
Λ22	Πληροφόρηση	-3	-2	-1	0	1	2	3
Λ23	Πορνογραφία	-3	-2	-1	0	1	2	3
Λ24	Προβλήματα	-3	-2	-1	0	1	2	3
Λ25	Στοργή	-3	-2	-1	0	1	2	3
Λ26	Σύγκρουση	-3	-2	-1	0	1	2	3
Λ27	Συναναστροφή	-3	-2	-1	0	1	2	3
Λ28	Τυχερά παιχνίδια	-3	-2	-1	0	1	2	3
Λ29	Φυγή	-3	-2	-1	0	1	2	3
Λ30	Ψέματα	-3	-2	-1	0	1	2	3

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Στη συγκεκριμένη έρευνα έλαβαν μέρος 581 νέοι, αγόρια και κορίτσια ηλικίας μεταξύ 18 και 35 ετών, εκ των οποίων 331 αγόρια και 250 κορίτσια.

Ο πίνακας δεδομένων που θα χρησιμοποιηθεί για τις αναλύσεις μετά των μετασχηματισμό που υπέστησαν οι αρχικές τιμές που αποδόθηκαν στις 30 λέξεις έχει ως εξής:

Πίνακας 2

Λ/Α	Αγχος	Αναγνωριση	Ανεπιθυμητοτητα	Ανοχη	Γνωση	Απόθεση	Αποσκέδαση	Εγκριση/αποψη	Εγγραψαι	Εικονικη	Παρέα	Πληροφορηση	Πορνογραφια	Προβληματισ	Σπουγη	Σύγκρουση	Συναισθημοφι	Τυχηρά
11	2	2	1	2	3	3	2	1	1	2	2	2	3	2	1	2	3	4	4	3
12	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
13	1	2	3	4	3	2	1	2	3	4	2	1	2	3	4	2	1	2	3	4
14	3	3	3	3	3	3	3	3	2	2	3	3	3	2	2	2	3	2	2	3
15	3	3	5	4	3	5	7	1	6	2	4	4	1	1	4	1	1	2	2	2
.....
1580	4	5	5	6	7	6	6	7	5	4	4	5	6	5	3	5	7	5	4	6
1581	4	5	1	4	5	6	7	7	4	7	6	5	7	7	4	5	6	7	7	7

Για την επεξεργασία των δεδομένων, η κλίμακα -3 έως +3 αντικαθίσταται με μία ισοδύναμη κλίμακα βαθμολογημένη από το 1 έως το 7, όπου το -3 της αρχικής αντιστοιχεί προφανώς στην τιμή 1 της νέας κλίμακας, ενώ το +3 στην τιμή 7.

Από τον πίνακα 2 δημιουργείται ο πίνακας 3 ο οποίος παρουσιάζει την γενική εικόνα με το πώς βαθμολόγησαν οι 581 νέοι τις 30 λέξεις.

Πίνακας 3: Οι βαθμολογίες των 30 λέξεων

ind	--1-	--2-	--3-	--4-	--5-	--6-	--7-		ind	--1-	--2-	--3-	--4-	--5-	--6-	--7-
Λ1	83	125	146	172	33	16	6		Λ16	86	74	136	140	104	29	12
Λ2	15	58	133	181	116	56	22		Λ17	87	64	135	153	95	37	10
Λ3	55	73	166	178	80	23	6		Λ18	81	79	125	150	97	39	10
Λ4	29	46	181	155	116	52	2		Λ19	84	69	125	140	108	40	15
Λ5	13	43	108	118	128	105	66		Λ20	20	44	84	163	131	74	65
Λ6	8	40	90	135	132	96	80		Λ21	8	37	100	135	133	98	70
Λ7	17	29	95	132	126	88	94		Λ22	25	39	96	144	124	79	74
Λ8	71	71	127	165	103	33	11		Λ23	74	85	121	158	90	35	18
Λ9	26	45	126	201	124	48	11		Λ24	50	90	145	147	99	44	6
Λ10	32	39	107	176	123	80	24		Λ25	21	52	106	189	107	67	39
Λ11	33	55	127	177	117	42	30		Λ26	46	72	142	183	97	33	8
Λ12	22	50	88	142	129	59	91		Λ27	26	55	114	138	137	80	31
Λ13	44	89	123	131	129	43	22		Λ28	35	69	158	165	80	40	34
Λ14	25	43	79	145	125	106	58		Λ29	39	86	152	177	61	37	29
Λ15	77	94	135	145	84	40	6		Λ30	76	88	135	182	43	33	24

Σημείωση: Την λέξη Αγχος (Λ1) 83 άτομα από τα 581 την βαθμολόγησαν με 1 (δηλαδή με -3)

Η ανάλυση του πίνακα 3 με το λογισμικό MAD παρουσιάζει τα εξής όσο αφορά τον τριδιάστατο παραγοντικό χώρο 1x2x3, ο οποίος ερμηνεύει το 92,52% της συνολικής πληροφόρησης που αντλείται από τον πίνακα 3.

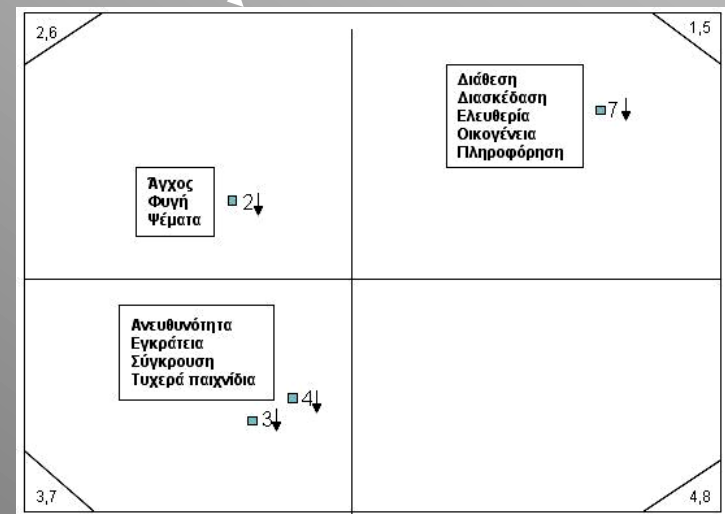
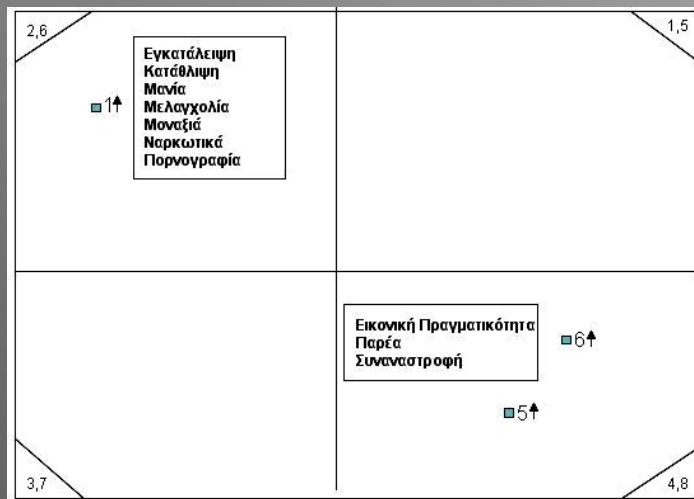
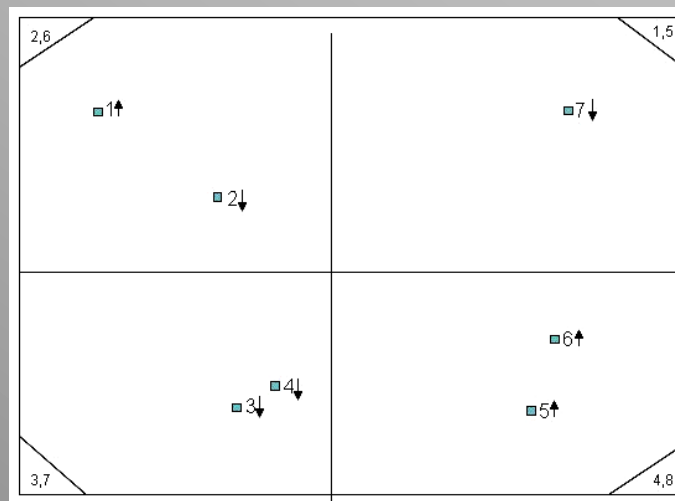
Πίνακας 7: Η κατανομή των 30 λέξεων στον παραγοντικό χώρο 1x2x3

Υποχώρος	1ος	2ος	3ος	4ος	5ος	6ος	7ος	8ος
Πλήθος	2	7	3	3	5	3	4	3
	Γνώση	Εγκατάλειψη	Ανοχή	Εικονική Πραγματικότητα	Διάθεση	Άγχος	Ανευθυνότητα	Αναγνώριση
	Επικοινωνία	Κατάθλιψη	Εξάρτηση	Παρέα	Διασκέδαση	Φυγή	Εγκράτεια	Εκκλησία
		Μανία	Προβλήματα	Συναναστροφή	Ελευθερία	Ψέματα	Σύγκρουση	Στοργή
		Μελαγχολία			Οικογένεια		Τυχερά Παιχνίδια	
		Μοναξιά			Πληροφόρηση			
		Ναρκωτικά						
		Πορνογραφία						

Πίνακας 7α: Η κατανομή των 7 βαθμών στον παραγοντικό χώρο 1x2x3

Υποχώρος	1ος	2ος	3ος	4ος	5ος	6ος	7ος	8ος
Βαθμοί		1 (-3)		5 (+1)	7 (+3)	2 (-2)	3 (-1)	
				6 (+2)			4 (0)	

Ο συνδυασμός των πινάκων 7 και 7α καταδεικνύει ότι οι επτά (7) λέξεις του 2ου υποχώρου (Εγκατάλειψη, Κατάθλιψη, Μανία, Μελαγχολία, Μοναξιά, Ναρκωτικά και Πορνογραφία) συνδέονται με την πλέον αρνητική βαθμολογία (-3), δείχνοντας την αποστροφή των νέων σε αυτές τις θλιβερές καταστάσεις που αναδύουν, ενώ οι πέντε (5) λέξεις του 5ου υποχώρου (Διάθεση, Διασκέδαση, Ελευθερία, Οικογένεια και Πληροφόρηση) συνδέονται με την πλέον θετική βαθμολογία (+3). Η αντιπαράθεση των δύο αυτών ομάδων λέξεων σημαίνει ότι οι νέοι Αγόρια και Κορίτσια έχουν αρχές και γερές βάσεις για να οικοδομήσουν μία σωστή κοινωνία.



Οι λέξεις του 1ου ,3ου και 8ου υποχώρου είναι «ορφανές» από βαθμούς επειδή στους υποχώρους δεν συγκαταλέγονται προφίλ βαθμών. Γι' αυτό προχωρούμε στην εφαρμογή της μεθόδου KARAP η οποία θα καλύψει αυτό το κενό.

Από τον πίνακα 2, με τη βοήθεια της στήλης που περιλαμβάνει τους κωδικούς του φύλου (1 ή 2), δημιουργείται ο συγκριτικός πίνακας 11 με αποτέλεσμα την δημιουργία του πίνακα 12 ο οποίος παρουσιάζει με ποιες λέξεις και σε ποιο βαθμό κυρίως επηρεάστηκαν τα αγόρια και τα κορίτσια

Πίνακας 11: Τμήμα του συγκριτικού πίνακα δεδομένων μεταξύ Αγοριών (ΛΑ1..ΛΑ30) και Κοριτσιών (ΛΚ1.. ΛΚ30)

IND	ΛΑ1	ΛΑ2	ΛΑ3	...	ΛΑ28	ΛΑ29	ΛΑ30	ΛΚ1	ΛΚ2	ΛΚ3	...	ΛΚ28	ΛΚ29	ΛΚ30
I1	2	2	1	...	4	4	3	2	2	2	...	4	3	2
I2	2	3	2	...	3	2	3	1	2	3	...	4	4	3
...
I250	4	3	4	...	4	4	4	4	5	1	...	7	7	7
I251	4	4	4	...	3	3	3	0	0	0	...	0	0	0
I252	2	7	1	...	5	4	6	0	0	0	...	0	0	0
...
I330	4	4	4	...	4	3	3	0	0	0	...	0	0	0
I331	4	5	5	...	5	4	6	0	0	0	...	0	0	0

Πίνακας 12: Τμήμα του πίνακα δεδομένων με τις συχνότητες των βαθμών που απέδωσαν τα Αγόρια και τα Κορίτσια στις 30 λέξεις

Πίνακας 13: Ταξινόμηση των δεδομένων του πίνακα 12 με την μέθοδο KARAP

ΑΓΟΡΙΑ						
-3	-2	-1	0	+1	+2	+3
Μανία	Κατάθλιψη	Ανοχή	Αναγνώριση		Διασκέδαση	Ελευθερία
Μελαγχολία	Πορνογραφία	Προβλήματα	Στοργή		Επικοινωνία	
Μοναξιά	Φυγή	Σύγκρουση			Οικογένεια	
Ναρκωτικά	Ψέματα	Τυχερά παιχνίδια			Πληροφόρηση	

ΚΟΡΙΤΣΙΑ						
-3	-2	-1	0	+1	+2	+3
	Άγχος	Ανευθυνότητα	Εγκράτεια	Εικονική πραγματικότητα	Γνώση	
		Εγκατάλειψη	Εκκλησία	Συναστροφή	Διάθεση	
		Εξάρτηση			Παρέα	

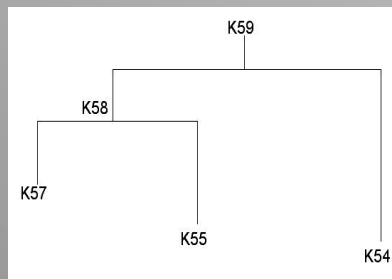
Πίνακας 16: Κατανομή των προφίλ των αγοριών και κοριτσιών που είναι πλησιέστερα στο προφίλ κάθε μιας από τις 30 λέξεις

1	2	3	4	5	6	7	8
ΛΕΞΕΙΣ	ΣΤΟ ΣΥΝΟΛΟ	%		ΑΓΟΡΙΑ	%	ΚΟΡΙΤΣΙΑ	%
Άγχος	14	2,41		10	3,02	4	1,6
Αναγνώριση	11	1,89		4	1,21	7	2,8
Ανευθυνότητα	34	5,85		26	7,85	8	3,2
Ανοχή	44	7,57		30	9,06	14	5,6
Γνώση	19	3,27		4	1,21	15	6
Διάθεση	29	4,99		11	3,32	18	7,2
Διασκέδαση	11	1,89		5	1,51	6	2,4
Εγκατάλειψη	25	4,3		16	4,83	9	3,6
Εγκράτεια	26	4,48		15	4,53	11	4,4
Εικονική πραγματικότητα	16	2,75		8	2,42	8	3,2
Εκκλησία	8	1,38		6	1,81	2	0,8
Ελευθερία	9	1,55		6	1,81	3	1,2
Εξάρτηση	18	3,1		10	3,02	8	3,2
Επικοινωνία	5	0,86		4	1,21	1	0,4
Κατάθλιψη	23	3,96		14	4,23	9	3,6
Μανία	19	3,27		10	3,02	9	3,6
Μελαγχολία	26	4,48		20	6,04	6	2,4
Μοναξιά	18	3,1		9	2,72	9	3,6
Ναρκωτικά	10	1,72		8	2,42	2	0,8
Οικογένεια	2	0,34		2	0,60	0	0
Παρέα	73	12,56		31	9,37	42	16,8
Πληροφόρηση	10	1,72		4	1,21	6	2,4
Πορνογραφία	12	2,07		9	2,72	3	1,2
Προβλήματα	24	4,13		14	4,23	10	4
Στοργή	1	0,17		1	0,30	0	0
Σύγκρουση	36	6,2		21	6,34	15	6
Συναναστροφή	6	1,03		3	0,91	3	1,2
Τυχικά παιχνίδια	23	3,96		14	4,23	9	3,6
Φυγή	16	2,75		10	3,02	6	2,4
Ψέματα	13	2,24		6	1,81	7	2,8
ΣΥΝΟΛΟ	581	100		331	100	250	100

ΤΟ ΨΥΧΟΓΡΑΦΗΜΑ ΤΩΝ 581 ΕΡΩΤΩΜΕΝΩΝ

Για να καταρτιστεί το πάζλ του ψυχογραφήματος των 581 ερωτηθέντων πρέπει πρώτα να εφαρμοστεί η μέθοδος VACOR στα δεδομένα του πίνακα 3, ώστε να προκύψει το δενδρόγραμμα των 30 λέξεων, από το οποίο θα εντοπιστούν οι κλάσεις με τα ιδιαίτερα χαρακτηριστικά της κάθε μιας, τα οποία θα αποτελούν το ψυχογράφημα των συμμετεχόντων σ' αυτήν.

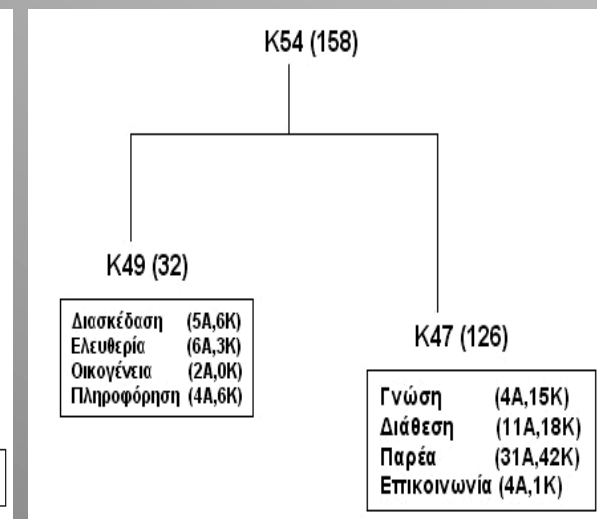
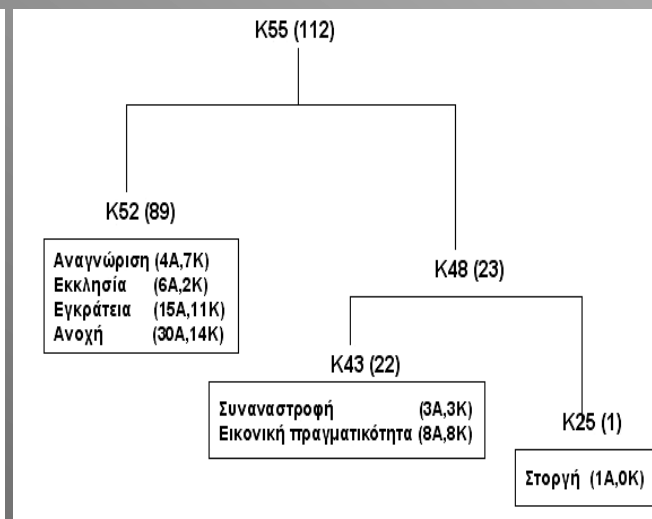
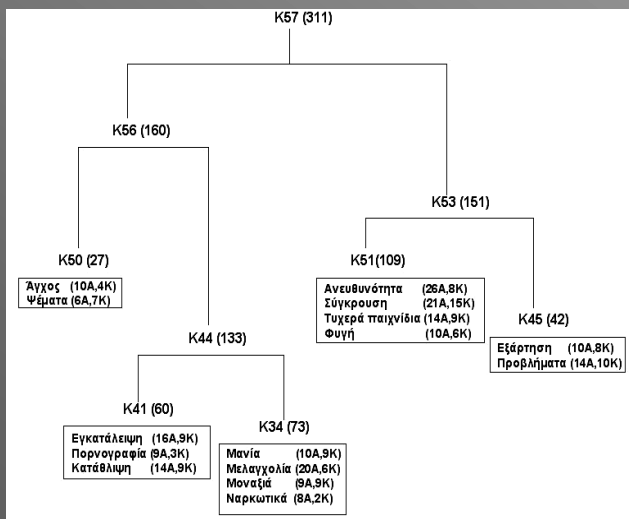
Η κλάση K59 των 30 λέξεων αρχικά διασπάται σε δύο κλάσεις την K58, και την K54, ενώ η κλάση K58 διασπάται σε δύο άλλες την K57 και την K55. Στη συνέχεια θα αναλυθούν οι τρεις αυτές κλάσεις χωριστά.



➤Όσον αφορά την κλάση K57

➤Όσον αφορά την κλάση K55

➤Όσον αφορά την κλάση K54



Αυτές οι πέντε κλάσεις αντιπροσωπεύουν τα πέντε ψυχολογικά προφίλ 331 ερωτώμενων οι οποίοι παρουσιάζουν **αρνητική** ψυχολογία.

Αυτές οι τρεις κλάσεις αντιπροσωπεύουν τα τρία ψυχολογικά προφίλ 112 ερωτώμενων οι οποίοι παρουσιάζουν **θετική** ψυχολογία.

Οι δύο αυτές κλάσεις που αντιπροσωπεύουν 158 ερωτώμενους οι οποίοι παρουσιάζουν πολύ **αισιόδοξη** ψυχολογία.

Ε) ΕΝΤΟΠΙΣΜΟΣ ΑΚΡΑΙΩΝ ΑΠΑΝΤΗΣΕΩΝ ΕΝΟΣ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟΥ

Με αυτή την εργασία παρουσιάζουμε την διαδικασία εντοπισμού ακραίων (ύποπτων) απαντήσεων ενός ερωτηματολογίου, οι οποίες δεν ανταποκρίνονται σε συμπεριφορές της πλειοψηφίας των ερωτώμενων

Κατά την ανάλυση του πίνακα αποστάσεων, όταν μελετάται ο παραγοντικός χώρος 1x2x3 με υψηλό ποσοστό ερμηνευτικής ικανότητας και εμφανίζεται το φαινόμενο κατά το οποίο κανένας από τους ερωτώμενους ΔΕΝ συνδέεται με καμία διαβάθμιση, αυτό σημαίνει απλά ότι οι συγκεκριμένοι ερωτώμενοι έχουν κάποια **ιδιαίτερη συμπεριφορά** διαφορετική από τους υπόλοιπους ερωτηθέντες, οπότε δικαίως οι απαντήσεις τους να θεωρούνται ως **ακραίες (ύποπτες) απαντήσεις**, οι οποίες χρήζουν ιδιαίτερης ερευνητικής μεταχείρισης, αφού σε αντίθετη περίπτωση η συμμετοχή τους στα συμπεράσματα, προκαλούν στρεβλές εκτιμήσεις για το υπό εξέταση θέμα.

Η έρευνα διεξήχθη από 1/10/2009 έως 30/10/2009 σε δείγμα 676 ατόμων και των δύο φύλων. Στην έρευνα έλαβαν μέρος 399 Άνδρες και 277 Γυναίκες.

1. Ποιος είναι ο κυριότερος παράγοντας αγοράς ενός κινητού τηλεφώνου (ΑΓΟ)

Η χαμηλή τιμή	[]	1
Η ωραία εμφάνιση	[]	2
Η υψηλή τεχνολογία	[]	3

2. Ποια είναι συνήθως η διάρκεια των συνομιλιών με το κινητό σου; (ΣΥΝ)

Συνήθως ολιγόλεπτες	[]	1
Περισσότερο από 10 λεπτά η κάθε μία	[]	2

3. Γνωρίζεις ότι η ηλεκτρομαγνητική ακτινοβολία του κινητού σου δεν πρέπει να ξεπερνά τη διάρκεια ενός εξαλέπτου σε μία συνομιλία κατά την διάρκεια ενός 24ώρου; (ΗΛΑ)

Όχι	[]	1
Ναι	[]	2

4. Γνωρίζεις ότι όλα τα κινητά δεν ακτινοβολούν το ίδιο λόγω κατασκευής τους; (ΑΚΤ)

Όχι	[]	1
Ναι	[]	2

5. Χρησιμοποιείς εξαρτήματα αποδέσμευσης των χεριών (hands free) και ακουστικών; (ΑΠΟ)

Όχι	[]	1
Ναι	[]	2

6. Ποια προληπτικά μέτρα από τα παρακάτω λαμβάνεις για την μείωση της ακτινοβολίας που δέχεσαι από την χρήση του κινητού σου; (ΠΡΟ)

Είμαι όσο το δυνατόν πιο σύντομος στις συνομιλίες	[]	1
Προμηθεύομαι κινητό με χαμηλό SAR	[]	2
Κανένα	[]	3

7. Ποια είναι η δική σου πεποίθηση για τα κινητά τηλέφωνα όσον αφορά την δημόσια υγεία των χρηστών (ΔΥΤ)

Δεν έχω άποψη	[]	1
Δεν υπάρχουν επιστημονικά στοιχεία για βλαβερές επιδράσεις	[]	2
Υπάρχει σημαντικός κίνδυνος	[]	3

8. Πραγματοποιείς ανακύκλωση του κινητού σου; (ΑΝΑ)

Όχι	[]	1
Ναι	[]	2

Πίνακας 1: Τμήμα του κωδικοποιημένου πίνακα των απαντήσεων

IND	ΑΓΟ	ΣΥΝ	ΗΛΑ	ΑΚΤ	ΑΠΟ	ΠΡΟ	ΔΥΓ	ΑΝΑ
I1	1	2	1	1	2	3	2	2
I2	3	2	2	2	1	2	3	1
I3	3	2	2	1	1	3	3	1
...
...
I676	3	1	1	1	2	2	1	2

Εφαρμόζοντας την μέθοδο KARAP έχουμε τα ακόλουθα αποτελέσματα στον τριδιάστατο παραγοντικό χώρο 1x2x3

Τοποθέτηση των 676 ατόμων και των 19 διαβαθμίσεων στους 8 υποχώρους του χώρου 1x2x3

A/A	1 ^{ος}	2 ^{ος}	3 ^{ος}	4 ^{ος}	5 ^{ος}	6 ^{ος}	7 ^{ος}	8 ^{ος}
ΠΑΗΘΟΣ	117	57	77	49	53	77	126	120
	I27	I4	I3	I1	I7	I2	I5	I10
	I33	I31	I8	I9	I19	I11	I6	I17
	I49	I56	I15	I16	I69	I12	I13	I18
	I52	I68	I20	I23	I82	I29	I14	I25
	I61	I80	I21	I24	I84	I32	I34	I26
	I103	I102	I22	I39	I94	I37	I40	I35
	I104	I105	I28	I41	I96	I50	I48	I36
	I108	I106	I30	I47	I98	I51	I64	I58
	I115	I107	I38	I53	I100	I54	I77	I62
	I122	I109	I42	I59	I173	I70	I79	I66

1 ^{ος}	2 ^{ος}	3 ^{ος}	4 ^{ος}	5 ^{ος}	6 ^{ος}	7 ^{ος}	8 ^{ος}
ΑΓΟ3	ΣΥΝ1	ΗΛΑ1	ΣΥΝ2	ΠΡΟ3	ΑΓΟ1	ΑΓΟ2	
ΗΛΑ2	ΠΡΟ1	ΑΚΤ1	ΑΚΤ2			ΔΥΓ2	
ΔΥΓ3	ΔΥΓ1	ΑΠΟ1	ΑΠΟ2			ΑΝΑ1	
ΑΝΑ2			ΠΡΟ2				

Από τον συνδυασμό των αποτελεσμάτων των πινάκων 6 και 7 προκύπτει ότι στον 8ο υποχώρο ΔΕΝ υπάρχει σύνδεση των 120 ερωτώμενων με καμία από τις 19 διαβαθμίσεις των 8 ερωτημάτων. Το συμπέρασμα αυτό ερμηνεύεται σε ποσοστό 81,4 % της πληροφορίας που αντλείται από τα δεδομένα του πίνακα $T_{dis}(676,19)$. Με βάση αυτή την πολύ ισχυρή παρατήρηση πρέπει να εντοπιστούν οι απαντήσεις των 120 ερωτώμενων και να κριθούν αν πράγματι είναι ακραίες, οπότε πρέπει να αφαιρεθούν, ώστε να προκύψει ένα δείγμα απαλλαγμένο από «**θορύβους**» που στρεβλώνουν το τελικό συμπέρασμα..

Ο πίνακας δεδομένων των 120 ατόμων

α/α	ΕΝΔΕΙΞΕΙΣ	ΑΓΟ	ΣΥΝ	ΗΛΑ	ΑΚΤ	ΑΠΟ	ΠΡΟ	ΔΥΓ	ΑΝΑ
1	I10	3	1	2	2	2	1	1	1
2	I17	2	1	2	1	2	1	3	1
3	I18	3	1	1	1	2	1	3	1
...
118	I668	3	1	1	2	1	1	1	1
119	I670	1	1	1	1	1	1	1	1
120	I676	3	1	1	1	2	2	1	2

Με βάση τα στοιχεία του πίνακα 8, παρατηρείται ότι η **πλειονότητα των 120 ερωτηθέντων** έχουν επιλέξει ως απάντηση την **πρώτη διαβάθμιση κάθε ερώτησης με κωδικό 1**, που σε ποσοστό είναι **κατά πολύ μεγαλύτερο** από εκείνο των υπόλοιπων ερωτηθέντων (πίνακας 9).

Εφαρμόζοντας τον έλεγχο ομοιότητας μεταξύ των ποσοστών των δύο δειγμάτων που αφορούν τους 120 ερωτηθέντες και τους υπόλοιπους 556, για κάθε πρώτη διαβάθμιση των 8 ερωτημάτων προκύπτουν τα εξής αποτελέσματα με βάση την z κατανομή, αφού σε κάθε περίπτωση ισχύει και για τα δύο δείγματα η συνθήκη $n_p > 5$ και $n_q > 5$

Πίνακας 9: Πίνακας ποσοστών των δύο δειγμάτων και η τιμή της z κατανομής που αντιστοιχεί στον έλεγχο ομοιογένειας κάθε διαβάθμισης.

Ένδειξη	ΑΓΟ1	ΣΥΝ1	ΗΛΑ1	ΑΚΤ1	ΑΠΟ1	ΠΡΟ1	ΔΥΓ1	ΑΝΑ1
Πλήθος	61	111	96	102	93	108	107	101
% στους 120	50,8	92,5	80	85	77,53	90	89,17	84,17
Πλήθος	63	224	208	163	181	99	120	265
% στους 556	11,33	40,29	37,41	29,32	32,6	17,81	21,58	47,66
Τιμή z	10,14	10,37	8,51	11,33	9,09	15,56	14,22	7,28

Από τον πίνακα 9 και τις τιμές της z κατανομής προκύπτει αβίαστα, σε επίπεδο σημαντικότητας 5%, ότι υφίσταται **σημαντική στατιστική διαφορά** για τον τρόπο που απάντησαν οι 120 ερωτώμενοι στο σύνολο των συγκεκριμένων διαβαθμίσεων των 8 ερωτημάτων, επομένως **δεν αποτελούν ενιαίο στατιστικό δείγμα** με εκείνο των 556 ερωτώμενων.

Αυτή η διαπίστωση οδηγεί στο να **αφαιρεθούν** από τον αρχικό πίνακα δεδομένων οι 120 ακραίες απαντήσεις, οπότε υφίσταται πλέον ένας **νέος πίνακας δεδομένων**, ο οποίος όταν αναλυθεί με την μέθοδο KARAP θα βγάλει περισσότερα ξεκάθαρα συμπεράσματα για το θέμα που διαπραγματεύεται.

Η νέα εικόνα της ανάλυσης των 556 ατόμων είναι η ακόλουθη με πιο ξεκάθαρα συμπεράσματα για την συμπεριφορά του ενιαίου πλέον δείγματος.

Πίνακας 10: Τμήμα του νέου πίνακα δεδομένων χωρίς τις 120 ακραίες απαντήσεις

α/α	ΕΝΔΕΙΞΕΙΣ	ΑΓΟ	ΣΥΝ	ΗΛΑ	ΑΚΤ	ΑΠΟ	ΠΡΟ	ΔΥΓ	ΑΝΑ
1	I1	1	2	1	1	2	3	2	2
2	I2	3	2	2	2	1	2	3	1
.
16	I19	2	1	2	2	1	2	3	1
17	I20	3	1	1	2	2	3	3	2
.
555	I674	3	1	1	1	2	1	2	1
556	I675	3	2	2	2	2	2	3	2

Εφαρμόζοντας στον πίνακα 10 την μέθοδο KARAP προκύπτουν τα ακόλουθα αποτελέσματα

Πίνακας 12: Η νέα ένταξη των 19 διαβαθμίσεων των 8 ερωτημάτων στους 8 υποχώρους κάνοντας χρήση τις νέες συντεταγμένες τους στο σύνολο των νέων παραγοντικών αξόνων

1 ^{ος}	2 ^{ος}	3 ^{ος}	4 ^{ος}	5 ^{ος}	6 ^{ος}	7 ^{ος}	8 ^{ος}
ΑΓΟ2	ΗΛΑ2	ΔΥΓ3	ΑΓΟ1	ΗΛΑ1	ΣΥΝ2	ΑΓΟ3	ΠΡΟ3
ΔΥΓ2	ΑΚΤ2		ΣΥΝ1	ΑΚΤ1			ΔΥΓ1
	ΑΠΟ2		ΠΡΟ1	ΑΠΟ1			
	ΠΡΟ2			ΑΝΑ1			
	ΑΝΑ2						

Πίνακας 13: Τμήμα από την νέα ένταξη των 556 ερωτηθέντων στους 8 υποχώρους κάνοντας χρήση τις νέες συντεταγμένες τους στο σύνολο των παραγοντικών αξόνων

⊕	Υποχώροι	1 ^{ος}	2 ^{ος}	3 ^{ος}	4 ^{ος}	5 ^{ος}	6 ^{ος}	7 ^{ος}	8 ^{ος}
	ΠΑΗΘΟΣ	55	37	81	43	47	77	71	145
		I21	I3	I7	I2	I5	I1	I19	I4
		I22	I8	I32	I12	I6	I16	I33	I11
		I28	I9	I50	I31	I13	I23	I103	I29
		I43	I39	I68	I70	I14	I24	I104	I40
		I44	I47	I69	I79	I15	I27	I105	I51
		I112	I49	I82	I80	I20	I30	I107	I54
		I164	I57	I84	I113	I34	I41	I108	I56
		I223	I60	I96	I162	I37	I42	I124	I81
		I236	I71	I98	I170	I38	I52	I125	I83

Ανάλυση με βάση τους υποχώρους

Με τον εντοπισμό των ερωτηθέντων σε κάθε υποχώρο του παραγοντικού χώρου 1x2x3 μπορούμε εφόσον το επιθυμούμε να αναλύσουμε για κάθε υποομάδα χωριστά τα δημογραφικά και κοινωνικοοικονομικά τους στοιχεία ταυτόχρονα με τα υπόλοιπα στοιχεία που συλλέχθηκαν κατά την έρευνα, έχοντας ενοποιήσει με ένα κωδικό το φύλο και την ηλικία των ερωτηθέντων για περισσότερο λεπτομερή συμπεράσματα που αφορούν τα δύο φύλα σε συνδυασμό με την ηλικία τους.

Με την ενέργεια της ενοποίησης σε ένα κωδικό για τα δύο φύλα και τις τρεις ηλικίες, δημιουργείται μια νέα μεταβλητή με έξι κωδικούς. Η κατανομή των 676 ερωτηθέντων, από τα οποία 399 είναι αγόρια και 277 κορίτσια έχει ως εξής:

Φύλο και Ηλικία	Κωδικός	Πλήθος
Αγόρια ηλικίας από 18 έως 20 ετών	Φ+Η1	57
Κορίτσια ηλικίας από 18 έως 20 ετών	Φ+Η2	24
Αγόρια ηλικίας από 21 έως 25 ετών	Φ+Η3	260
Κορίτσια ηλικίας από 21 έως 25 ετών	Φ+Η4	159
Αγόρια ηλικίας από 26 έως 35 ετών	Φ+Η5	82
Κορίτσια ηλικίας από 26 έως 35 ετών	Φ+Η6	94
ΣΥΝΟΛΟ		676

Αφαιρώντας τα 120 ερωτηματολόγια αναλύουμε τον νέο πίνακα δεδομένων με την μέθοδο KARAP , μελετάμε χωριστά τις 8 υποομάδες ατόμων που έχει η κάθε μια ομοιογένεια ως προς το ΦΥΛΟ και την ΗΛΙΚΙΑ , ταυτόχρονα με συγκεκριμένες διαβαθμίσεις των 6 νέων μεταβλητών με ταυτότητα Φ+Η, οι οποίες χαρακτηρίζουν τις υποομάδες

Π.χ για την 1^η υποομάδα των 55 ατόμων έχουμε τα ακόλουθα αποτελέσματα

Πίνακας 19: Η ένταξη των 25 πλέον διαβαθμίσεων στους 8 υποχώρους

Υποχώρος	1ος	2ος	3ος	4ος	5ος	6ος	7ος	8ος
	Φ+Η5	ΣΠΟ2	ΟΙΚ2	Φ+Η2	Φ+Η1	ΚΑΤ1	Φ+Η3	ΣΠΟ1
	ΚΑΤ2		ΔΥΓ2	Φ+Η6	Φ+Η4	ΠΡΟ3	ΔΙΑ1	ΟΙΚ4
	ΔΙΑ2		ΑΝΑ1	ΣΠΟ3	ΔΥΓ3	ΑΝΑ2	ΟΙΚ1	
	ΟΙΚ3			ΚΑΤ3			ΠΡΟ1	
							ΔΥΓ1	

Πίνακας 20: Η ένταξη των 55 ερωτηθέντων στους 8 υποχώρους

Υποχώρος	1ος	2ος	3ος	4ος	5ος	6ος	7ος	8ος
Πλήθος	6	6	9	7	11	4	6	6

ΕΦΑΡΜΟΓΕΣ ΤΩΝ ΜΕΘΟΔΩΝ ΤΗΣ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ

Περιλαμβάνονται στο έργο **ΑΝΘΟΛΟΓΙΑ** του συγγραφέα

A) Εφαρμογές κλασσικών μεθόδων ανάλυσης δεδομένων

Ανάλυση ενός πίνακα συμπτώσεων
Προσδιορισμός του πλήθους των παραγόντων που είναι σημαντικοί για την ερμηνεία ενός πίνακα δεδομένων
Επιλογή τοποθεσίας για δημιουργία έκθεσης επίπλων
Ανάλυση τεχνικών χαρακτηριστικών 15 πολυτελών αυτοκινήτων
Προσδιορισμός του ρυθμού κατανάλωσης CD ROM παιχνιδιών
Ταξινόμηση των τίτλων CD ROM ως προς την ζήτηση που παρουσιάζουν
Ταξινόμηση βάσει του συντελεστή συσχέτισης
Ένταξη ανώνυμων ερωτώμενων σε γνωστές ομάδες με την μέθοδο της Διακριτικής Ανάλυσης
Τμηματοποίηση της αγοράς
Τακτοποίηση νέων ερωτώμενων σε προκαθορισμένες ομάδες
Τοποθέτηση υπαλλήλων βάσει της αποτελεσματικότητάς τους
Διακριτική ικανότητα τεσσάρων παραγόντων που διαμορφώνουν την ποιότητα ενός κρασιού
Ανάδυση συναισθημάτων. Η μέθοδος της Σημειομετρίας
Ασύμμετρη ταξινόμηση. Η μέθοδος της Συνεπαγωγικής Στατιστικής
Επεξεργασία τριδιάστατων πινάκων $I \times J \times T$. Η περίπτωση των εμπορικών συναλλαγών της Ελλάδος με τα κράτη οκτώ γεωγραφικών περιοχών του πλανήτη την περίοδο 2002-2004
Ανάλυση μιας ακολουθίας k συνεχόμενων πινάκων συμπτώσεων

B) Νέα εργαλεία στη διερεύνηση πολυδιάστατων δεδομένων

Μεθοδολογία της διερεύνησης δεδομένων με βάση την παραγοντική ανάλυση πολυδιάστατων δεδομένων

Η μέθοδος KARAP ως εργαλείο εξόρυξης δεδομένων

Αξιολόγηση της Ανιούσας Ιεραρχικής Ταξινόμησης. Μέθοδος BENKAR

Η εφαρμογή της μεθόδου KARAP ως εργαλείο σχεδιασμού διαφημιστικής εκστρατείας ενός τουριστικού προορισμού

Χαρτογράφηση της ψυχολογίας των νέων χρηστών του internet. VACOR vs KARAP

Βελτίωση της ερμηνευτικής ικανότητας ενός λογικού πίνακα 0-1 με την χρήση της μεθόδου KARAP

Εντοπισμός ακραίων απαντήσεων ενός ερωτηματολογίου

Δημιουργία διαφημιστικού μηνύματος με την μέθοδο KARAP

Γ) Κωδικοποίηση δεδομένων βάσει ενός λογικού πίνακα 0-1

Ανάλυση ενός λογικού πίνακα 0-1

Κωδικοποίηση και ανάλυση μιας λογικής διαδικασίας

Μια τεχνική βελτίωσης της πληροφορίας που παρέχει ένας πίνακας BURT

Ειδική επεξεργασία ερωτηματολογίου με ερωτήματα πολλαπλής απάντησης

Οι απόψεις των σπουδαστών του Αλεξάνδρειου Τ.Ε.Ι Θεσσαλονίκης του τμήματος εμπορίας και διαφήμισης (Μάρκετινγκ), όσον αφορά στα ναρκωτικά κατά την χρονική περίοδο Απριλίου- Μαΐου 1998

Δ) Μέθοδοι έρευνας κοινής γνώμης

Μέθοδοι έρευνας κοινής γνώμης και αγοράς με πολυδιάστατες μεθόδους
Η στάση των καταναλωτριών στην εμφάνιση ενός νέου απορρυπαντικού πιάτων
Μία νέα προσέγγιση επεξεργασίας ερωτηματολογίων έρευνας αγοράς με βάση τις αρχές της σημειομετρίας
Προσδιορισμός καταναλωτικών προτιμήσεων με τη χρήση της Ιεραρχικής Ανάλυσης
Ιεράρχηση της προτίμησης έξι Ελληνικών νησιών ως τουριστικός προσδιορισμός
Οι απόψεις των σπουδαστών του Αλεξάνδρειου Τ.Ε.Ι. Θεσσαλονίκης του Τμήματος Εμπορίας και Διαφήμισης (Μάρκετινγκ), όσον αφορά στα ναρκωτικά κατά την χρονική περίοδο Απριλίου- Μαΐου 1998.

Ε) Κλασικές μέθοδοι αξιολόγησης

Αξιολόγηση των προτιμήσεων των Κυπρίων για την επιλογή του τόπου των διακοπών τους
Αξιολόγηση τριών κριτηρίων για την αγορά ενός προϊόντος
Συγκριτική μελέτη του βαθμού ικανοποίησης των χρηστών από τις υπηρεσίες που προσφέρονται κατά τις εφημερίες και τις προγραμματισμένες επισκέψεις στα εξωτερικά ιατρεία ενός Νοσοκομείου
Συγκριτική αξιολόγηση k ομάδων βάσει p κριτηρίων με την χρήση ενός πίνακα αποστάσεων
Η αξιολόγηση του μαθήματος της Ανάλυσης Δεδομέ-νων από τους φοιτητές του Ε! εξαμήνου κατά το εαρινό εξάμηνο του ακαδημαϊκού έτους 2004-05
Αξιολόγηση των υπηρεσιών που προσφέρει το ΑΤ.Ε.Ι.Θ στους φοιτητές

ΣΤ) Πολιτικό Μάρκετινγκ και Ανάλυση Δεδομένων

Προβλήματα με τα οποία δεν ασχολούνται επαρκώς τα τρία μεγάλα κόμματα της Βουλής των Ελλήνων σύμφωνα με τις απόψεις των ψηφοφόρων τους

Σύγκριση πολιτικών κειμένων των κομμάτων του ΠΑΣΟΚ της Ν.Δ και του ΚΚΕ πριν τις εκλογές του 2004

Η ιδεολογική πλατφόρμα επικοινωνίας της ΝΔ, του ΠΑΣΟΚ και του ΚΚΕ με τους ψηφοφόρους

Αξιολόγηση της δημοφιλίας ενός ηγέτη κόμματος από τους ψηφοφόρους του

Η) Ανάλυση χρηματοοικονομικών δεδομένων

Πρόβλεψη της μελλοντικής συμπεριφοράς (τάσης) μιας χρονοσειράς

Η διαχρονική εξέλιξη του δείκτη των λιανικών πωλήσεων κατά την χρονική περίοδο 1998-2002

Συνθετική προσέγγιση: Μία νέα μεθοδολογία προσδιορισμού της Χρηματιστηριακής Αγοράς ως θερμοδυναμικό σύστημα για την κατάρτιση του Φερέγγου Χαρτοφυλακίου με μεθόδους Ανάλυσης Δεδομένων

Προσδιορισμός του Φερέγγου χαρτοφυλακίου

Αξιολόγηση Αμοιβαίων Κεφαλαίων. Η μέθοδος του Επιμερισμού του κινδύνου

Θ) Ανάλυση Δεδομένων και νέες τεχνολογίες

Στην ενότητα αυτή θα παρουσιαστούν εργασίες του Δρος Δημητρίου Καραπιστόλη σε συνεργασία με άλλους ερευνητές στην Ελληνική και Αγγλική γλώσσα, οι οποίες παρουσιάστηκαν σε διάφορα συνέδρια και δημοσιεύθηκαν σε Ελληνικά και ξένα περιοδικά διεθνούς κύρους.

Οι εργασίες αυτές παρουσιάζουν τις ξεχωριστές δυνατότητες συνεργασίας των μεθόδων της Ανάλυσης Δεδομένων χρησιμοποιώντας νέες τεχνολογίες της επιστήμης της πληροφορικής και όχι μόνο.

Factor Data Analysis and Econophysics: application in market segmentation. Dimitrios Karapistolis and George Stalidis. *International Conference on Econophysics Kavala 2-3 June 2011* Journal of Engineering Science and Technology Review 4 (3) (2011) 256 – 260 Special Issue on Econophysics

«**Multidimensional Data Analysis and Knowledge Management Technologies for Tourist Decision Support**», Stalidis G. and Karapistolis D. Proc of 3rd International Conf Quantitative and Qualitative Methodologies in the Economic and Administrative Sciences (QMEAS) Athens, 23-24 May, 2013, pp 373-379 ISBN: 978-960-98739-4-9

Σύγκριση Τεχνικών Μηχανικής Μάθησης βάσει των αποτελεσμάτων της Ανιούσας Ιεραρχικής Ταξινόμησης. Δρ. Ειρήνη Δημ. Καραπιστόλη. Τετράδια Ανάλυσης Δεδομένων, 17, σελ.14-28.

«**Knowledge discovery and computerized reasoning to assist tourist destination marketing**» Stalidis G. and Karapistolis D. *International Journal on Strategic Innovative Marketing* , Vol.01,pp103-119DOI:10.1556/IJSIM.01.02.004

«**Marketing decision support using Artificial Intelligence and Knowledge Modelling: application to tourist destination management**». Stalidis G., Karapistolis D. and Vafeiadis A. 3rd Intl Conference on Strategic Innovative Marketing, Madrid 1-4 September 2014.Procedia - Social and Behavioral Sciences (2015) pp.106-113 DOI:10.1016/J.SBSPRO.2015.01.1180

«**Tourist Destination Marketing Supported by Electronic Capitalization of Knowledge**» Stalidis G. and Karapistolis D. *Procedia - Social and Behavioral Sciences* 148 (2014) 110 – 118

«**Codage et analyse d' une epreuve logique**» (1986), *Les Cahiers de l' Analyse des Donnees*» Vol.IX 1986 pp 482-486 (Μεταπτυχιακή εργασία)

ΒΙΒΛΙΑ ΤΟΥ ΣΥΓΓΡΑΦΕΑ

Δρος ΔΗΜΗΤΡΙΟΥ ΝΙΚΟΛΑΟΥ ΚΑΡΑΠΙΣΤΟΛΗ

1. ΦΕΡΕΓΓΥΟ ΧΑΡΤΟΦΥΛΑΚΙΟ Συνθετική Προσέγγιση της εξέλιξης των χρηματιστηριακών τιμών των μετοχών Εκδόσεις Ανικούλα 1999

Το παρόν βιβλίο παρουσιάζει για πρώτη φορά στο αναγνωστικό κοινό, μία νέα μέθοδο διαχείρισης χαρτοφυλακίων, χρησιμοποιώντας μοντέρνες αντιλήψεις που αφορούν όχι μόνο τις στατιστικές μεθόδους ανάλυσης, αλλά κυρίως την φιλοσοφία που διέπει τον τρόπο αντιμετώπισης της εξέλιξης των χρηματιστηριακών τιμών των μετοχών. Ειδικότερα η χρηματιστηριακή αγορά αντιμετωπίζεται ως θερμοδυναμικό σύστημα το οποίο διέπεται από χαοτική δυναμική η οποία οφείλεται στην προσδοκία των εμπλεκομένων μερών να αποκομίσουν χρηματικά οφέλη. Μια τέτοια θεώρηση οδηγεί τον αναλυτή να υιοθετήσει, για τα τεκτανόμενα στη χρηματιστηριακή αγορά, μια αντίληψη τελεολογικού χαρακτήρα, η οποία τον βοηθά να κατανοήσει τις δομικές διαφοροποιήσεις που υφίσταται το συγκεκριμένο σύστημα κατά την εξελικτική του πορεία. Η νέα μεθοδολογία, η οποία ονομάζεται Συνθετική Προσέγγιση, αποτελεί ένα επιστημονικό σύστημα επεξεργασίας της διαθέσιμης πληροφορίας, το οποίο επιδιώκει να προσδιορίσει την κλιμακούμενη φερεγγυότητα των επιχειρήσεων που συμμετέχουν στη χρηματιστηριακή αγορά, βάσει της οποίας καταρτίζεται το Φερέγγυο Χαρτοφυλάκιο με μεθόδους πολυπαραγοντικής ανάλυσης δεδομένων.



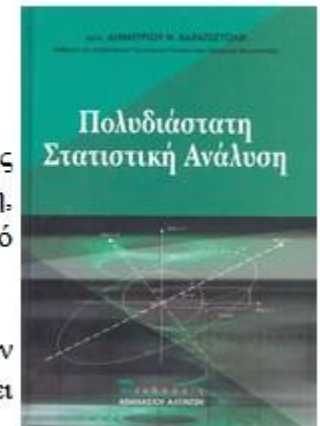
2. ΠΟΛΥΔΙΑΣΤΑΤΗ ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ Εκδόσεις ' . Αλτιντζή 2011

Το παρόν βιβλίο «Πολυδιάστατη Στατιστική Ανάλυση» παρουσιάζει αναλυτικά τα εξής θέματα:

α) Την αναγκαιότητα των μεθόδων της ανάλυσης δεδομένων στη μελέτη των οικονομικών φαινομένων.

β) την περιγραφή των πολυδιάστατων μεθόδων: την παραγοντική ανάλυση των αντιστοιχιών, την ανάλυση σε κύριες συνιστώσες, την ανιούσα ιεραρχική ταξινόμηση, την ανάλυση των τάξεων, την παραγοντική διακριτική ανάλυση, την ιεραρχική ανάλυση και την συνεπαγωγική στατιστική η οποία παρουσιάζεται για πρώτη φορά στο Ελληνικό κοινό.

γ) Μέσα από τις σελίδες του βιβλίου αναδύεται η σύνδεση μεταξύ αρχών και νόμων της Φυσικής και των αρχών που διέπουν τις μεθόδους της Πολυπαραγοντικής Στατιστικής Ανάλυσης η οποία είναι εμφανής, όταν ανατρέξει κανείς στα κεφάλαια II και X, αποτελώντας πολύτιμο βοήθημα.



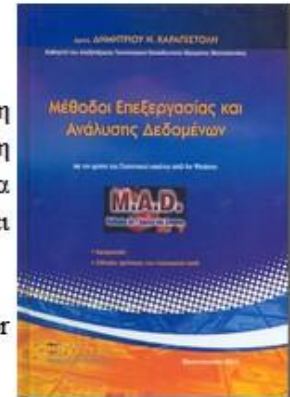
3. ΜΕΘΟΔΟΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΚΑΙ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ

Με την χρήση του στατιστικού πακέτου MAD for Windows. Εκδόσεις Α. Αλτιντζή 2011

Το παρόν βιβλίο αποτελείται από δύο μέρη:

Στο 1ο μέρος παρουσιάζονται αναλυτικά εφαρμογές όλων των πολυδιάστατων παραγοντικών μεθόδων, όπως η Παραγοντική ανάλυση των Αντιστοιχιών, η Αιούσα Ιεραρχική Ταξινόμηση, η Ανάλυση σε Κύριες Συνιστώσες, η Παραγοντική Διακριτική Ανάλυση, η Στατιστική Συνεπαγωγή, η Ιεραρχική Ανάλυση, ώστε ο αναγνώστης να κατανοήσει σε βάθος την χρησιμότητα των μεθόδων αυτών, αλλά και το ποιες μεθόδους πρέπει να εφαρμόσει ανάλογα με τα δεδομένα που θέλει να επεξεργαστεί.

Στο 2ο μέρος παρουσιάζονται αναλυτικά και με παραδείγματα οι οδηγίες χρήσης του Στατιστικού πακέτου MAD for Windows, που υλοποιεί τις προαναφερόμενες μεθόδους πολυδιάστατης στατιστικής ανάλυσης



4. ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗΣ ΜΕΤΟΧΩΝ ΚΑΙ ΑΜΟΙΒΑΙΩΝ ΚΕΦΑΛΑΙΩΝ. Εκδόσεις Α. Αλτιντζή 2013

Το παρόν βιβλίο έχει ως στόχο να εισαγάγει κάθε ενδιαφερόμενο επενδυτή στις χρηματιστηριακές έννοιες και στις μεθόδους της διαχείρισης χρηματιστηριακών χαρτοφυλακίων τις οποίες χρησιμοποιούν οι μνημόνοι του χώρου της χρηματιστηριακής αγοράς, ώστε να έχει την δυνατότητα μόνος του να αξιολογεί τις πληροφορίες που συλλέγει και να διαμορφώνει τη δική του επενδυτική στρατηγική.

Στο 1ο κεφάλαιο γίνεται αναφορά στο τι πρεσβεύει η θεωρία της αποτελεσματικότητας της αγοράς

Στο 2ο κεφάλαιο αναπτύσσεται περιληπτικά το τι πραγματεύεται η θεμελιακή ανάλυση

Στο 3ο κεφάλαιο περιγράφονται οι διαδικασίες πρόβλεψης με βάση τα πιο ευρέως διαδεδομένα εργαλεία της Τεχνικής ανάλυσης

Στο 4ο κεφάλαιο αναπτύσσεται η σύγχρονη θεωρία χαρτοφυλακίου

Στο 5ο κεφάλαιο παρουσιάζεται η νέα θεωρία διαχείρισης χαρτοφυλακίων «Συνθετική Προσέγγιση» και το προϊόν αυτής το Φερέγγυο Χαρτοφυλάκιο

Στο 6ο κεφάλαιο αναπτύσσεται η νέα μεθοδολογία αξιολόγησης των Αμοιβαίων Κεφαλαίων «Μέθοδος του επιμερισμού των αποδόσεων» η οποία στηρίζεται στην εκτίμηση του μάνατζμεντ της ΑΕΔΑΚ που το διαχειρίζεται. Η εργασία αυτή αποτελεί μονογραφία και παγκόσμια πρωτοτυπία της οποίας τα πνευματικά δικαιώματα είναι καταχωρημένα με συμβολαιογραφική πράξη αριθμός 1973, της 8/10/1997 του συμβολαιογράφου Θεσσαλονίκης Εμμανουήλ Δ. Σαραγιώτη



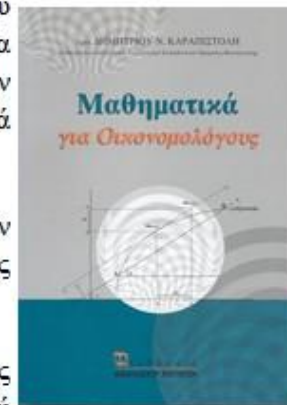
5. ΜΑΘΗΜΑΤΙΚΑ ΓΙΑ ΟΙΚΟΝΟΜΟΛΟΓΟΥΣ Εκδόσεις Α. Αλπιντζής 2011

Το βιβλίο αυτό περιλαμβάνει την ύλη που προβλέπεται στο αναλυτικό πρόγραμμα του μαθήματος Οικονομικά μαθηματικά, αντικείμενο το οποίο διδάσκεται στα Τμήματα των Σχολών Διοίκησης και Οικονομίας των Τεχνολογικών Εκπαιδευτικών Ιδρυμάτων της χώρας (ΤΕΙ).

Ο αντικειμενικός σκοπός του βιβλίου είναι να εισαγάγει τον φοιτητή στις βασικές έννοιες του αντικείμενου που εξετάζεται, χωρίς να υπεισέρχεται σε λεπτομερή ανάλυση, παραθέτοντας όμως πολλαπλά παραδείγματα, τα οποία θα τον καταστήσουν ικανό να αποκτήσει την απαιτούμενη αναλυτική σκέψη στη μελέτη των καθημερινών προβλημάτων που θα συναντήσει στην επαγγελματική του σταδιοδρομία. Η ενασχόληση με τα Οικονομικά Μαθηματικά προϋποθέτει βέβαια γενικότερες γνώσεις της Μαθηματικής επιστήμης.

Στο Α! μέρος αναπτύσσονται οι κατ' εξοχήν έννοιες των οικονομικών, δίνονται κάποιες εφαρμογές των μαθηματικών στην Οικονομική επιστήμη, αφού προηγουμένως έχουν αναφερθεί θεμελιώδεις έννοιες της οικονομικής ανάλυσης. Στη συνέχεια αναπτύσσονται ο απλός τόκος, ο ανατοκισμός, οι Ράντες και τα Δάνεια.

Στο Β! μέρος, γίνεται ανασκόπηση των βασικών εννοιών της επιστήμης των μαθηματικών, όπως θέματα πρακτικής αριθμητικής, Άλγεβρας, καθώς και στοιχεία αναλυτικής γεωμετρίας. Έτσι ο ασχολούμενος με την Οικονομική επιστήμη, διαθέτει ένα συνοπτικό μνημόνιο με τις πλέον απαραίτητες έννοιες, το οποίο βέβαια σε καμιά περίπτωση δεν υποκαθιστά τα βιβλία που έχουν γραφεί πάνω σ' αυτά τα αντικείμενα, αλλά βοηθούν απλώς να θυμηθεί ο αναγνώστης λησμονημένες έννοιες, τις οποίες αν θέλει να τις μελετήσει σε βάθος θα πρέπει να ανατρέξει στα εξειδικευμένα συγγράμματα του είδους.



6. ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ Εκδόσεις Α. Αλπιντζής 2012

Σκοπός του παρόντος βιβλίου, του οποίου η περιεχόμενη ύλη δεν καλύπτει ολόκληρη τη στατιστική μεθοδολογία, είναι να μεταδώσει στους φοιτητές των Τεχνολογικών Εκπαιδευτικών Ιδρυμάτων (ΤΕΙ), αλλά και σε οποιονδήποτε άλλο ενδιαφερόμενο, τις γενικές αρχές πάνω στις οποίες στηρίζεται η στατιστική μελέτη ποσοτικών κυρίως μεγεθών στα πλαίσια κάθε επιστημονικής έρευνας.

Για την καλύτερη κατανόηση των στατιστικών μεθόδων που αναπτύσσονται στα διάφορα κεφάλαια, παρατίθενται σε κάθε ένα από αυτά σχετικά παραδείγματα, τα οποία ικανοποιούν τον περιγραφικό χαρακτήρα της ανάλυσης ενός οικονομικού μεγέθους, καθόσον η εις βάθος ανάλυσή του απαιτεί εξειδικευμένες γνώσεις οι οποίες ανήκουν σε γνωστικά αντικείμενα, όπως λ.χ. η οικονομετρία, η συστηματική προσέγγιση, η ανάλυση δεδομένων και η χασοπή θεωρία.



Το παρόν βιβλίο παρουσιάζει για πρώτη φορά στο αναγνωστικό κοινό (στην Αγγλική), μία νέα μέθοδο διαχείρισης αρχείων, χρησιμοποιώντας μοντέρνες αντιλήψεις που αφορούν όχι μόνο τις στατιστικές μεθόδους ανάλυσης, αλλά κυρίως την φιλοσοφία που διέπει τον τρόπο αντιμετώπισης των δεδομένων.



New research tools with
multidimensional statistic
analysis methods

LAMBERT

Δρ. ΔΗΜΗΤΡΙΟΣ Ν. ΚΑΡΑΠΙΣΤΟΛΗΣ

ΘΕΩΡΙΑ ΤΟΥ ΕΠΙΜΕΡΙΣΜΟΥ ΤΩΝ ΑΠΟΔΟΣΕΩΝ ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ Α/Κ

ΕΚΔΟΣΕΙΣ hitechPC

Θεσσαλονίκη 2017

Δρος ΔΗΜΗΤΡΙΟΥ Νικ. ΚΑΡΑΠΙΣΤΟΛΗ

Αξιολόγηση Προσωπικού

Εκτίμηση της οργανωσιακής κουλτούρας της επιχείρησης

Με μεθόδους Πολυδιάστατης Στατιστικής Ανάλυσης Δεδομένων



Θεωρία-Μέθοδοι / Το project DANSTEY



Εκδόσεις
Λιβανασίου Αλτιντζή

Δρος ΔΗΜΗΤΡΙΟΥ Νίκ. ΚΑΡΑΠΙΣΤΟΛΗ

ΕΓΧΕΙΡΙΔΙΟ ΙΕΡΑΡΧΙΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ

ΜΕ ΘΕΜΕΛΙΟ ΤΗ ΓΑΛΛΙΚΗ ΣΧΟΛΗ
ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ

ΘΕΩΡΙΑ - ΕΦΑΡΜΟΓΕΣ



ΘΕΣΣΑΛΟΝΙΚΗ 2024

Το λογισμικό MAD περιλαμβάνει:

-Δημιουργία αρχείων κατάλληλων προς επεξεργασία με τις διαφορετικές μεθόδους, ανάλυσης του στατιστικού πακέτου.

-Λήψη αρχείων από EXCEL και SPSS

-Διαδικασίες επεξεργασίας μονοδιάστατων και πολυδιάστατων πινάκων,

-Κλασικές στατιστικές αναλύσεις, όπως Στατιστικούς ελέγχους, Ανάλυση διακύμανσης, Συσχέτιση, Παλινδρόμηση – Απλή και πολλαπλή, Χρονολογικές σειρές, Χαοτικές σειρές, Μη παραμετρικούς ελέγχους, Στατιστικούς πίνακες

-Μεθόδους ανάλυσης δεδομένων όπως Παραγοντική Ανάλυση των Αντιστοιχιών, Ανάλυση σε Κύριες Συνιστώσες, Ανάλυση Τάξεων, Ανιούσα Ιεραρχική Ταξινόμηση, Διακριτική Παραγοντική Ανάλυση, Διακρίνουσα Ανάλυση, -Συνεπαγωγική Στατιστική και Ιεραρχική Ανάλυση.

-Τις μεθόδους KARAP, BENKAR, ΣΥΝΘΕΤΗ ΙΕΡΑΡΧΙΚΗ ΤΑΞΙΝΟΜΗΣΗ, ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΚΛΑΣΕΩΝ ΤΗΣ ΑΝΙΟΥΣΑΣ ΙΕΡΑΡΧΙΚΗΣ ΤΑΞΙΝΟΜΗΣΗΣ

-Πολλές άλλες διαδικασίες οι οποίες παρουσιάζονται στην ηλεκτρονική διεύθυνση www.pylimad.gr

ΕΥΧΑΡΙΣΤΟΥΜΕ
για το ενδιαφέρον σας

Τηλέφωνα επικοινωνίας: 6977868954

Δρ. ΔΗΜΗΤΡΙΟΣ ΚΑΡΑΠΙΣΤΟΛΗΣ