

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΕΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ
ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ
ΤΟΜΕΑΣ ΓΕΩΛΟΓΙΑΣ
ΕΡΓΑΣΤΗΡΙΟ ΓΕΩΛΟΓΙΑΣ - ΠΑΛΑΙΟΝΤΟΛΟΓΙΑΣ

ΒΑΣΙΚΕΣ ΜΕΘΟΔΟΙ ΠΟΛΥΠΑΡΑΓΟΝΤΙΚΗΣ ΑΝΑΛΥΣΗΣ.
ΠΑΡΑΔΕΙΓΜΑΤΑ ΣΕ ΔΕΔΟΜΕΝΑ ΠΑΛΑΙΟΝΤΟΛΟΓΙΑΣ
ΘΗΛΑΣΤΙΚΩΝ.

Διπλωματική Εργασία
ΙΩΑΝΝΗ ΣΑΡΡΗ

Επιβλέπων Καθηγητής Γεώργιος Δ. Κουφός

Θεσσαλονίκη 2005

ΠΕΡΙΕΧΟΜΕΝΑ

- I. Πρόλογος
- II. Θεωρητική ανάλυση των μεθόδων
 - 1. Ανάλυση Κύριων Συνιστωσών
 - 2. Ανάλυση Ομάδων
 - 3. Διακρίνουσα Ανάλυση
- III. Εφαρμογές
 - 1. Το δείγμα
 - i. Περιγραφή δείγματος
 - ii. Επεξεργασία δείγματος
 - 2. Hadjidimono
 - 3. Πικέρμι
 - 4. *H. mediterraneum*
 - 5. *H. brachypus*
- IV. Σύνοψη – Συμπεράσματα
- V. Παράρτημα
- VI. Βιβλιογραφία

I. Πρόλογος

Η παρούσα μελέτη εντάσσεται στα πλαίσια του υποχρεωτικού μαθήματος «Διπλωματική εργασία» του Η' εξαμήνου του Τμήματος Γεωλογίας του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης. Σκοπός της εργασίας είναι η θεωρητική ανάλυση κάποιων βασικών πολυπαραγοντικών μεθόδων και η κατανόησή τους σε τέτοιο βαθμό ώστε να χρησιμοποιηθούν ως εργαλείο για την επίλυση κάποιων προβλημάτων στην Παλαιοντολογία. Η επιλογή του θέματος της εργασίας έγινε με γνώμονα την τροπή που έχει πάρει η παλαιοντολογική έρευνα τις τελευταίες δεκαετίες με την εφαρμογή ποσοτικής ανάλυσης. Μέσω της επιστήμης της βιολογίας, η ανάλυση δεδομένων έχει εισβάλει στην παλαιοντολογία ενισχύοντας κυρίως την παραδοσιακή ποιοτική ανάλυση και όχι αντικαθιστώντας την.

Αρχικά θα γίνει η παρουσίαση των μεθόδων αναφέροντας τον σκοπό της κάθε μεθόδου αλλά και πως θα επιτευχθεί αυτός, κατανοώντας έτσι το τι εκφράζουν τα αποτελέσματα που προκύπτουν. Κατόπιν θα εφαρμοστούν αυτές οι μέθοδοι σε δείγματα απολιθωμένων ιπποειδών του γένους *Hipparion* ηλικίας Άνω Μειοκαίνου των απολιθωματοφόρων θέσεων Πικέρμι (Ελλάδα) και Hadjidimono (Βουλγαρία). Τέλος θα γίνει η ερμηνεία των αποτελεσμάτων καθώς και η αξιολόγησή τους.

Η εργασία έγινε στο Εργαστήριο Γεωλογίας και Παλαιοντολογίας του Α.Π.Θ. υπό την επίβλεψη του καθηγητού κ. Γεωργίου Δ. Κουφού. Τα α δεδομένα των μετρήσεων από το Πικέρμι μου δόθηκαν από τον κ. Γ. Δ. Κουφό, ενώ του Hadjidimono από την κ. L. Christova.

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Γ. Δ. Κουφό για την ανάθεση και παρακολούθηση αυτής της διπλωματικής εργασίας. Ευχαριστώ επίσης τον Δρ. Δ. Κωστόπουλο για την βοήθεια του στην επεξεργασία και ανάλυση των δεδομένων. Ακόμα θα ήθελα να ευχαριστήσω τον υποψήφιο Δρ. Μ. Στύλλα καθώς και τον Δρ. Κ. Βουδούρη για την βοήθειά τους πάνω σε θέματα Στατιστικής. Τέλος ευχαριστώ την κ. L. Christova για την παραχώρηση των μετρήσεων από τα ιπάρια του Hadjidimono.

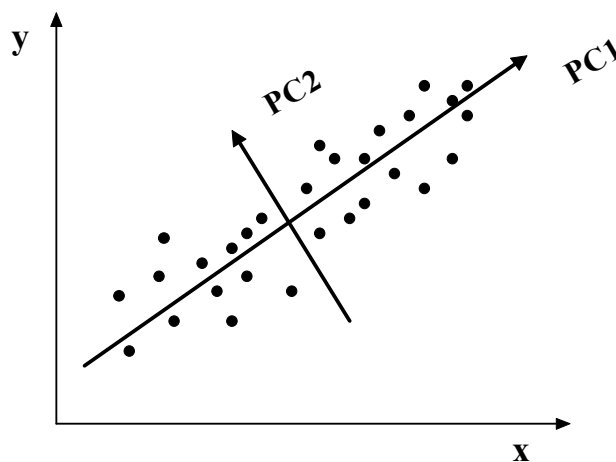
II. Θεωρητική Ανάλυση των Μεθόδων

II.1 Ανάλυση Κύριων Συνιστωσών (Principal Components Analysis)

Ο στόχος της Ανάλυσης Κύριων Συνιστωσών (PCA) είναι η προβολή ενός πολυπαραγοντικού δείγματος σε λιγότερες διαστάσεις. Ουσιαστικά υπάρχει μια μείωση των μεταβλητών του δείγματος (συνήθως σε 2 ή 3) καθιστώντας εφικτή την απεικόνιση των παρατηρήσεων. Αυτή η μείωση γίνεται με γνώμονα την κατά το δυνατόν διατήρηση της διακύμανσης των δεδομένων και αυτό επιτυγχάνεται με την εύρεση λιγότερων διαστάσεων, κατά μήκος όμως των οποίων εμφανίζεται η μεγαλύτερη διακύμανση του

δείγματος. Οι καινούριες μεταβλητές που είναι γραμμικές συναρτήσεις των αρχικών ονομάζονται **κύριες συνιστώσες** (Principal Components ή απλώς PC). Οι κύριες συνιστώσες μπορούν να θεωρηθούν ως μεταβλητές, οι οποίες έχουν βιολογική σημασία.

Για να γίνει αντιληπτό το πώς λειτουργεί η PCA, έστω ένα διμεταβλητό σύστημα του οποίου οι δύο μεταβλητές σχετίζονται γραμμικά. Σε μια τέτοια περίπτωση η μέγιστη (αν όχι όλη) διακύμανση του δείγματος εμφανίζεται κατά μήκος της ευθείας ελαχίστων τετραγώνων. Έτσι η κύρια συνιστώσα με τη μεγαλύτερη ποσοστιαία συμμετοχή στην ολική διακύμανση θα βρίσκεται κατά μήκος αυτής της ευθείας. Αυτό γίνεται εύκολα αντιληπτό από το σχήμα 1.



Σχήμα-1. Διάγραμμα δύο μεταβλητών ενός δείγματος οι οποίες σχετίζονται γραμμικά. Το PC1 είναι κατά μήκος της ευθείας ελαχίστων τετραγώνων εκφράζοντας σχεδόν όλη τη διακύμανση του δείγματος, ενώ το PC2 είναι κάθετο σε αυτό και συμμετέχει μόνο κατά ένα μικρό ποσοστό. Η διαφορά της σχετικής συμμετοχής του καθενός στην ολική διακύμανση φαίνεται και από τα διαφορετικά μήκη (το PC1 έχει μεγαλύτερο μήκος).

Στην PCA οι κύριες συνιστώσες είναι ουσιαστικά τα ιδιοδιανύσματα [παράρτημα A.1] ενός πίνακα διακύμανσης – συνδιακύμανσης ή συσχέτισης [παράρτημα A.2]. Συνοπτικά η διαδικασία που ακολουθείται είναι η παρακάτω:

Έστω ένας πίνακας ενός δείγματος m μεταβλητών x_i , του οποίου οι σειρές αντιπροσωπεύουν τα αντικείμενα και οι στήλες τις μεταβλητές. Βρίσκουμε τον πίνακα διακύμανσης – συνδιακύμανσης [s^2] ο οποίος είναι διαστάσεων $m \times m$. Από αυτόν το πίνακα θα βρούμε m ιδιοδιανύσματα και m ιδιοτιμές (παράρτημα A.1), που θα ορίζουν τα στοιχεία ενός ελλειψοειδούς m διαστάσεων. Τα ιδιοδιανύσματα θα δίνουν τη διεύθυνση των κύριων αξόνων του ελλειψοειδούς, ενώ οι ιδιοτιμές το μήκος των ημιαξόνων του καθώς και το ποσοστό της ολικής διακύμανσης που ερμηνεύεται από αυτούς.

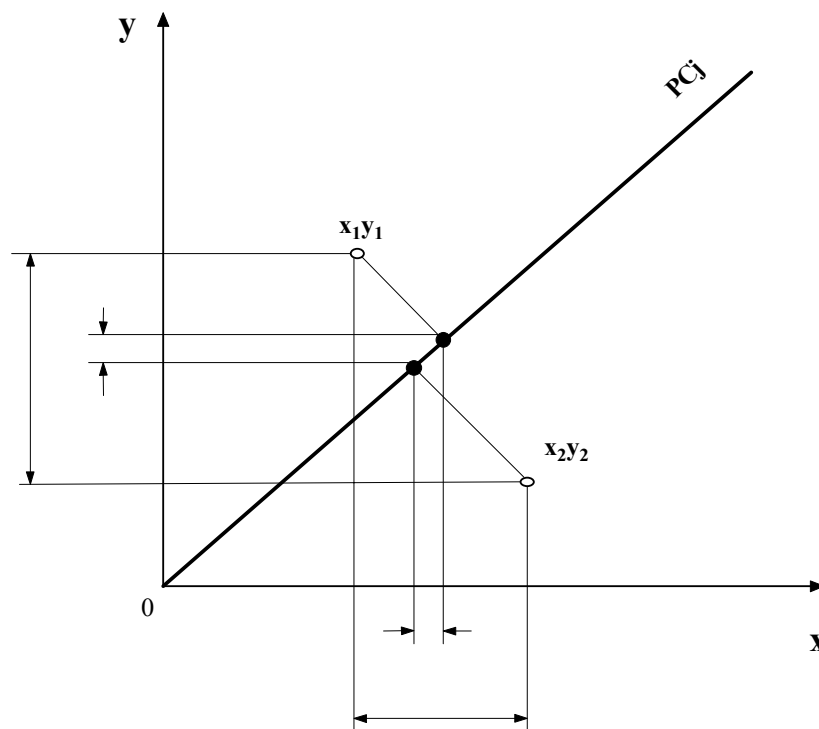
Η ολική διακύμανση ενός δείγματος ορίζεται ως το άθροισμα των επιμέρους διακυμάνσεων των μεταβλητών. Εφόσον το άθροισμα των αξόνων των κύριων συνιστωσών δίνει την ολική διακύμανση και οι ιδιοτιμές τα μήκη των ημιαξόνων των

κύριων συνιστωσών (δηλαδή τις διακυμάνσεις τους), αν διαιρεθεί η κάθε μία ιδιοτιμή με την ολική διακύμανση θα μας δώσει την ποσοστιαία συμμετοχή της αντίστοιχης κύριας συνιστώσας στην ολική διακύμανση.

Κάθε αρχική παρατήρηση μπορεί να μετατραπεί σε τιμή κάποιας κύριας συνιστώσας προβάλλοντας τις τιμές των μεταβλητών για τη συγκεκριμένη παρατήρηση πάνω στον άξονα αυτής της κύριας συνιστώσας. Σε μια γενικευμένη περίπτωση m μεταβλητών x_i , η τιμή της κύριας συνιστώσας j για το αντικείμενο k θα δίνεται από τον τύπο:

$$PC_{jk} = \sum_{i=1}^m a_{ij} x_{ik}$$

όπου a_{ij} είναι το **φορτίο** (loading) της μεταβλητής x_i για την κύρια συνιστώσα j . Τα φορτία είναι συντελεστές οι οποίοι εκφράζουν τη σχετική συμμετοχή κάθε αρχικής μεταβλητής στην κύρια συνιστώσα (PC). Γραφικά αυτή η προβολή φαίνεται στο σχήμα 2.



Σχήμα-2. Στο σχήμα έχουμε δύο μεταβλητές (x και y) και δυο αρχικές παρατηρήσεις (ο) x_1y_1 και x_2y_2 οι οποίες προβάλλονται πάνω στην j κύρια συνιστώσα (•). Ένα πολύ σημαντικό συμπέρασμα που εξάγεται από αυτό το σχήμα είναι ότι όταν αυτές οι τιμές της κύριας συνιστώσας προβληθούν με τη σειρά τους πάνω στους άξονες των αρχικών μεταβλητών εμφανίζουν μικρότερη διακύμανση από αυτή των αρχικών. Αυτό το έλλειμμα πληροφορίας καλύπτεται από τις υπόλοιπες κύριες συνιστώσες.

Τα παραπάνω μπορούν να εκφραστούν με την μορφή πινάκων ως εξής:

$$[X][U]=[S^R]$$

όπου $[S^R]$ είναι ο πίνακας $n \times m$ των τιμών των κύριων συνιστωσών, $[X]$ ο πίνακας $n \times m$ των αρχικών παρατηρήσεων και $[U]$ ο τετραγωνικός πίνακας m στηλών που περιέχει τα φορτία.

Σε ένα δείγμα του οποίου οι δυο μεταβλητές παρουσιάζουν μεγάλη συσχέτιση, η πρώτη κύρια συνιστώσα θα έχει μεγάλο μήκος απεικονίζοντας σχεδόν όλη τη διακύμανση του δείγματος, ενώ η δεύτερη θα είναι με ελάχιστον μήκος, ένδειξη του ότι δεν συμβάλει σημαντικά στη διακύμανση. Στην αντίθετη περίπτωση, όπου οι μεταβλητές του δείγματος δεν σχετίζονται καθόλου, οι κύριες συνιστώσες (όντας ίσου μήκους) ορίζουν ένα κύκλο.

Ένα σημαντικό ερώτημα που προκύπτει είναι το εξής: «πόσες κύριες συνιστώσες χρειάζονται για την εξαγωγή ενός ασφαλούς συμπεράσματος;». Ο Jolliffe (1986) πρότεινε ως απάντηση μια *ad hoc* διαδικασία στην οποία χρησιμοποιούνται οι κύριες συνιστώσες ως και την l^* εφαρμόζοντας τον τύπο:

$$l^* = 0.7(\sum_{i=1}^m l_i) / m$$

όπου το l_i είναι οι ιδιοτιμές του πίνακα συσχέτισης ή διακύμανσης – συνδιακύμανσης. Ο ίδιος ερευνητής αναφέρει ότι η διαδικασία αυτή φαίνεται να λειτουργεί καλύτερα με πίνακα συσχέτισης. Στην πραγματικότητα ο αριθμός των κύριων συνιστωσών που χρησιμοποιούνται εξαρτάται από τον κάθε ερευνητή.

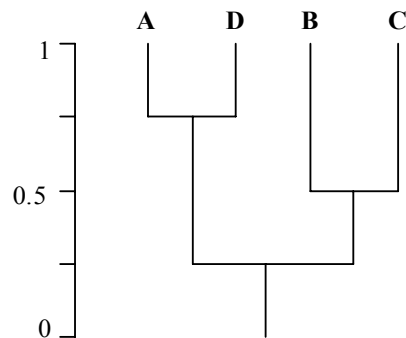
Η PCA είναι από τις πιο σημαντικές πολυπαραγοντικές μεθόδους και εφαρμόζεται ευρέως στην παλαιοντολογία. Είναι μία πολύ αποτελεσματική μέθοδος για την εκτίμηση αλλά και περιγραφή της μεταβλητότητας ανάμεσα σε είδη καθώς και μέσα στο ίδιο είδος και εφαρμόζεται για την επίλυση φυλογενετικών προβλημάτων σε ταξινομικό επίπεδο είδους. Επίσης εφαρμόζεται στην μορφολειτουργική ανάλυση, ενώ είναι ένα από τα σημαντικότερα εργαλεία της παλαιοοικολογίας.

II.2 Ανάλυση Ομάδων (Cluster analysis)

Η Ανάλυση Ομάδων έχει ως στόχο την τοποθέτηση παρατηρήσεων σε περισσότερο ή λιγότερο ομοιογενείς ομάδες οι οποίες θα είναι ευδιάκριτες. Ουσιαστικά η μεθοδολογία αναφέρεται σε ένα σύνολο μεθόδων ομαδοποίησης παρατηρήσεων. Στη παρούσα εργασία θα αναλυθούν μόνο οι ιεραρχικές τεχνικές οι οποίες και θα χρησιμοποιηθούν παρακάτω.

Στην ιεραρχική ομαδοποίηση συνδέονται αρχικά οι παρόμοιες παρατηρήσεις και ακολούθως προσαρτώνται στα σύνολα οι πιο παρεμφερείς σε αυτά. Αυτό επιτυγχάνεται με τη δημιουργία πίνακα ομοιοτήτων των παρατηρήσεων ανά ζεύγη. Από αυτό τον πίνακα, ο οποίος υπολογίζεται κάθε φορά που δημιουργείται μια ομάδα, εξάγεται το δενδρόγραμμα (σχήμα 3), που απεικονίζει τόσο τις ομάδες όσο και τις μεταξύ τους σχέσεις.

Για την δημιουργία ενός πίνακα ομοιοτήτων προτείνονται αρκετοί συντελεστές από τους οποίους σημαντικότεροι είναι ο *συντελεστής συσχέτισης* r_{ij} καθώς και μια τυποποιημένη *ευκλείδεια απόσταση* d_{ij} . Για την χρησιμοποίηση της ευκλείδειας απόστασης θα πρέπει ο πίνακας δεδομένων να τυποποιηθεί [Παράρτημα Γ.1] ώστε να συμμετέχουν όσο το δυνατόν λιγότερο οι παρατηρήσεις μεγάλου μεγέθους. Παρακάτω περιγράφεται συνοπτικά η τεχνική *weighted pair-group* με αριθμητικούς μέσους.



Σχήμα-3. Το δενδρόγραμμα των παρατηρήσεων A, B, C και D όπου τα A και D ορίζουν μία ομάδα με συντελεστή ομοιότητας 0.75, τα B και C αποτελούν μια άλλη ομάδα έχοντας συντελεστή ομοιότητας 0.5 ενώ και οι δύο ομάδες συνδέονται με συντελεστή 0.25.

Έστω ένας πίνακας δείγματος $n \times m$ (n παρατηρήσεις και m μεταβλητές). Υπολογίζουμε την συσχέτιση των μετρήσεων ανάμεσα σε όλα τα ζεύγη παρατηρήσεων δημιουργώντας έναν συμμετρικό πίνακα $n \times n$. Οποιοδήποτε στοιχείο (c_{ij}) αυτού του πίνακα εκφράζει την ομοιότητα του αντικειμένου i με το j . Έτσι για ένα δείγμα με $n=4$ προκύπτει ο πίνακας του σχήματος 4.

Τα αντικείμενα τα οποία εμφανίζουν αμοιβαία τις υψηλότερες συσχετίσεις, ορίζουν τις πρώτες ομάδες. Δυο αντικείμενα i και j εμφανίζουν αμοιβαία τις υψηλότερες συσχετίσεις όταν οι συντελεστές c_{ij} και c_{ji} έχουν την υψηλότερη τιμή ο καθένας στη στήλη του. Κατόπιν πρέπει να υπολογιστεί ξανά ο πίνακας συσχετίσεων θέτοντας αυτή τη φορά όμως τα αντικείμενα που περιέχονται σε μια ομάδα ως ένα. Αυτός ο υπολογισμός του νέου πίνακα ομοιοτήτων γίνεται με διάφορους τρόπους. Στην συγκεκριμένη τεχνική για την εύρεση της ομοιότητας ενός εξωτερικού αντικειμένου με μια ομάδα υπολογίζεται απλώς ο αριθμητικός μέσος των ομοιοτήτων του εξωτερικού αντικειμένου με τα στοιχεία της ομάδας.

	A	B	C	D
A	1	c_{AB}	c_{AC}	c_{AD}
B	c_{BA}	1	c_{BC}	c_{BD}
C	c_{CA}	c_{CB}	1	c_{CD}
D	c_{DA}	c_{DB}	c_{DC}	1

Σχήμα-4. Πίνακας συσχετίσεων όπου c_{ij} είναι ο συντελεστής ομοιότητας της i με την j παρατήρηση.

Ακολουθώντας αυτή τη μεθοδολογία θα προσθέτονται συνεχώς νέες παρατηρήσεις σε υπάρχουσες ομάδες, δημιουργώντας έτσι κάθε φορά ομάδες μικρότερου συντελεστή ομοιότητας (άρα και μικρότερης ομοιογένειας), έτσι θα προκύψει τελικά ένα *δενδρόγραμμα* της μορφής του σχήματος 3.

Με αυτή τη διαδικασία, δηλαδή θεωρώντας τα αντικείμενα μιας ομάδας ως ένα, δημιουργείται ένα παραμορφωμένο *δενδρόγραμμα* του οποίου η παραμόρφωση αυξάνεται στα διαδοχικά επίπεδα ομαδοποίησης. Για το λόγο αυτό υπάρχουν αρκετές τεχνικές που προσπαθούν να λύσουν αυτό το πρόβλημα έχοντας όμως άλλα μειονεκτήματα.

Η μέθοδος *single linkage clustering* η οποία ομαδοποιεί με βάση τη μέγιστη ομοιότητα του εξωτερικού αντικειμένου με ένα οποιοδήποτε στοιχείο της ομάδας, είναι κατάλληλη για υψηλά ταξινομικά επίπεδα. Υπάρχουν μέθοδοι που βασίζονται στη θεώρηση όλων των παρατηρήσεων ως ισοβαρών, αυτό όμως έχει ως συνέπεια το εξωτερικό αντικείμενο να επηρεάζει πολύ τα αποτελέσματα. Αυτή η σημαντική επιρροή των εξωτερικών στοιχείων αποφεύγεται στις *μη σταθμικές* μεθόδους (*unweighted method*), όπου οι ομάδες συμμετέχουν ποσοστιαία σύμφωνα με τον αριθμό των στοιχείων τους, αποφεύγοντας την σημαντική επιρροή των τελευταίων στοιχείων που εισάγονται στην ομάδα.

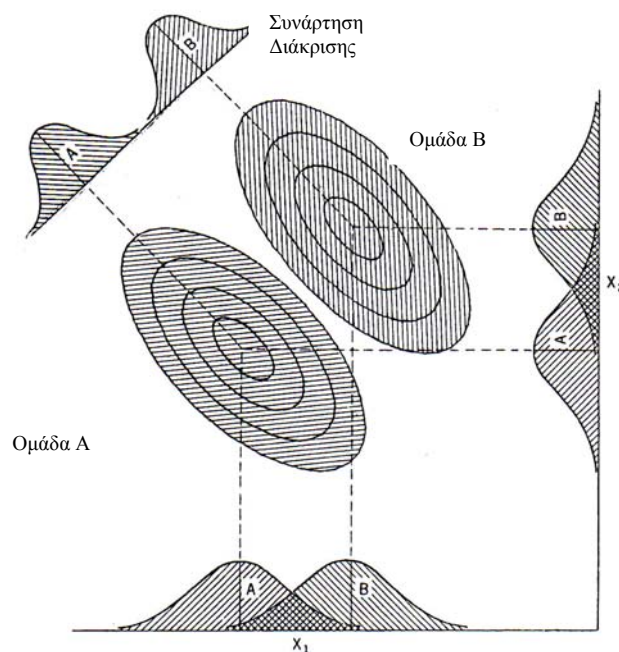
Δυστυχώς δεν υπάρχει κάποια μέθοδος που να θεωρείται η καλύτερη αν και οι *weighted pair-group* μέθοδοι φαίνεται ότι έχουν καλύτερα αποτελέσματα. Δυσκολία στην επιλογή της κατάλληλης μεθόδου έρχεται να προσθέσει και το γεγονός ότι δεν υπάρχουν στατιστικοί έλεγχοι για τις μεθόδους της ιεραρχικής ταξινόμησης.

II.3 Διακρίνουσα Ανάλυση (Discriminant analysis)

Σκοπός της *Διακρίνουσας Ανάλυσης* είναι η προβολή ενός πολυδιάστατου συστήματος σε μία διάσταση με τέτοιο τρόπο ώστε να μεγιστοποιείται η διαφοροποίηση ανάμεσα σε δύο *ήδη* καθορισμένα σύνολα. Αυτή η διάσταση της μέγιστης διαφοροποίησης εκφράζεται με μία γραμμική σχέση των μεταβλητών, τη *διακρίνουσα συνάρτηση*.

Η συνάρτηση διάκρισης είναι πολύ σημαντική για το λόγο του ότι μπορεί να χρησιμοποιηθεί στην τοποθέτηση ενός νέου δείγματος σε μία από τις δύο ομάδες. Μία απλή γραμμική συνάρτηση διάκρισης μετατρέπει ένα σύνολο μετρήσεων σε μια τιμή η οποία βρίσκεται κατά μήκος μιας ευθείας που ορίζεται από αυτή τη συνάρτηση. Σε αυτή την ευθεία υπάρχει μια κρίσιμη τιμή εκατέρωθεν της οποίας τοποθετούνται οι παρατηρήσεις των δύο ομάδων. Έτσι βρίσκοντας για ένα νέο δείγμα την τιμή που προέρχεται από τη συνάρτηση διάκρισης μπορούμε να το τοποθετήσουμε σε μια από τις δύο ομάδες. Με αυτό τον τρόπο ένα πολυπαραγοντικό πρόβλημα ανάγεται σε μονοδιάστατο και επιλύεται ευκολότερα.

Η ανάλυση διακρίνουσας συνάρτησης βρίσκει τον τρόπο μετατροπής που δίνει την ελάχιστη αναλογία της διαφοράς ανάμεσα στους πολυπαραγοντικούς μέσους όρους προς αυτήν των διακυμάνσεων των δύο δειγμάτων. Αυτό σημαίνει ότι σε έναν δισδιάστατο χώρο αν οι δυο ομάδες είναι σύνολα σημείων τότε πρέπει να βρούμε την ευθεία κατά μήκος της οποίας οι δυο ομάδες έχουν την μέγιστη διαφορά και ταυτόχρονα ελάχιστο εύρος (η κάθε μία) (σχήμα 5).



Σχήμα-5. Διάγραμμα δύο διμεταβλητών κατανομών που εμφανίζουν αλληλεπικάλυψη μεταξύ τους κατά μήκος των μεταβλητών x_1 και x_2 . Οι δυο αυτές ομάδες διακρίνονται καλά, αφού προβληθούν πάνω στην ευθεία που εκφράζεται από την συνάρτηση διάκρισης.

Για την εύρεση της διακρίνουσας συνάρτησης πρέπει να επιλυθεί η εξίσωση:

$$[S_p^{-2}] \cdot [\lambda] = [D]$$

όπου $[s_p^2]$ είναι ένας πίνακας $m \times m$ συγχωνευμένων διακυμάνσεων και συνδιακυμάνσεων m μεταβλητών (παράρτημα Β.1). Οι συντελεστές της συνάρτησης διάκρισης είναι τα στοιχεία του πίνακα λ , ενώ το δεξιό σκέλος της σχέσης είναι ένας πίνακας m διαφορών ανάμεσα στις μέσες τιμές των δύο ομάδων (παράρτημα Β.2).

Η συνάρτηση διάκρισης είναι της μορφής:

$$R = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_m x_m$$

όπου x είναι οι μεταβλητές και λ οι συντελεστές. Οι συντελεστές είναι αυτοί οι αριθμοί που θα καθορίσουν το κατά πόσο συνεισφέρει κάθε μεταβλητή στην τιμή που θα πάρει η εξίσωση, η οποία με τη σειρά της θα καθορίσει τη θέση του νέου δείγματος σε ένα από τα δύο σύνολα. Έχοντας υπόψη αυτό, οι μεταβλητές που δεν συνεισφέρουν αρκετά, δηλαδή δεν καθορίζουν το αποτέλεσμα, μπορούν να απορριφθούν.

Η επιτυχής διάκριση ελέγχεται με το ποσοστό των παρατηρήσεων που είναι τοποθετημένες επιτυχώς κατά μήκος του άξονα διαχωρισμού.

III. Εφαρμογές

III.1 Τα Απολιθωμένα Δείγματα

i. Περιγραφή δειγμάτων

Τα δείγματα που θα μελετηθούν προέρχονται από δύο απολιθωματοφόρες θέσεις του Α. μειοκαίνου των Βαλκανίων, το Πικέρμι που βρίσκεται και το Hadjidimono που βρίσκεται στη Ν. Βουλγαρία. Ο κύριος όγκος των ευρημάτων που προέρχονται από το Πικέρμι στεγάζεται στο Μουσείο Φυσικής Ιστορίας του Λονδίνου και του Παρισιού, ενώ το υλικό από το Hadjidimono βρίσκεται στο Assevonograd, περίπου 5Km από την Φιλιπούπολη στη Βουλγαρία. Οι απολιθωματοφόρες θέσεις χαρακτηρίζονται από μεγάλο αριθμό ευρημάτων μεταξύ των οποίων και δύο είδη ιππαρίου το *H. mediterraneum* και το *H. brachypus*. Ο μεγάλος αριθμός των δειγμάτων που αποδίδονται στα δύο αυτά είδη ιππαρίου κρίνονται επαρκή για μια πρώτη ποσοτική επεξεργασία των δεδομένων.

Για την εφαρμογή της Πολυπαραγοντικής ανάλυσης χρησιμοποιήθηκαν μόνο οι μετρήσεις που αφορούν το τρίτο μετακαρπικό (από και στο εξής MCIII), καθώς θεωρείται οστό που μπορεί να παρουσιάζει μορφολογική διαφοροποίηση από είδος σε είδος ιππαρίου (Vlachou & Koufos, 2002) και επιπλέον βρίσκεται σε ιδιαίτερη αφθονία, οπότε μπορεί να αποτελέσει ένα καλό στατιστικό δείγμα.

Τα δεδομένα προέρχονται από το αρχείο μετρήσεων του καθηγητή Γ. Δ. Κουφού (Πικέρμι) και της διδάκτορος L. Christova, University of Sofia (Βουλγαρία). Στους παρακάτω πίνακες (1 έως 4) δίνονται οι μετρήσεις των δειγμάτων ανά περιοχή και ανά είδος, όπου οι γραμμές κάθε πίνακα αντιπροσωπεύουν τα δείγματα και οι στήλες τις μεταβλητές, οι οποίες απεικονίζονται στο σχήμα 6.

	1	3	4	5	6	7	8	10	11	12	14
19391	229,5	23,5	21,5	41	26	33,5	12,2	35,4		28,9	
19383	218,5	23,8	22,8	37,4	28,5	32,5	10,5	33	34	28,5	25,2
19389	219	26,5	26,5	21,9				36,5		28,5	
19377	229,5	25,5	23,4					33,8	35,2		25
19406	233	26,3	22,4					36			27
19381	242,5	24,5	28,3	44	27,5	37	10,5	39,5	39,8	32	28,5
19427	239,5	26,8	23	42	26,8	34,5	10	37,8	36,8	30,5	27
19487	215,5	26,5	21,8	41	27	34,8	12		36,4		26
19461	213			40,7	24,8	34,2	11,5				
19539	224	24,6	22,5	39,2	27,4			34			
19538	217,7	26,1	21,5	41,2	27,4			39	37	31	27,5
19532	225,5	26,2	24,5	42	28,1	34	12,2	37,3	37,3	30,6	26,1
19540	221	22	20,2	38,5			10,9	32	34,5	27,5	25
19548	222	25,3	23	40,5	26	34,8	9,8		35,3	28	25,2
19501	214,6	26,5	22,2					34,5	33,8	29	26
19523	231,3	23	20,8					35,3	34,4	29	26,3
19525	229,5	25,3	22,5	41	27,3	34,7	10,2	34,4	35,5	30	26,6
17052	212,5	27	22	40,7	27	35,8	9,6	37,6	37,1	30,4	27,2
17100	231,6	24,6	22,9	36,5	28	35		35,4	34,7	27,5	25,2
17200	231,1	23,4	22,5	41,3	26,3	35,3	10,7	33,5	34,8	27,8	23,3
17222	226	24,3	22	38,6	26,5	32,5		35,3	35,3	29	26,2
19486	206,7	26,8	22,6					36		29	27,2

Πίνακας-1. *H. mediterraneum*, Hadjidimono, Βουλγαρία. Μετρήσεις του ΜСΠΙ (για τις μετρήσεις βλέπε σχ. 6)

	1	3	4	5	6	7	8	10	11	12	14
19401	194,5	29,5	20,22	40,8	27	35,7	9,5		35,5	39,6	
19382	211	32	23	42	27			41,5			
19400	210	30,5	23	43	29,1	38	11,8	42,5	40,5	32,3	27
19388	215,5	32,8	23	44,2	31	38,5	10,5			35	29,8
19386	213,5	33,5	25	47,5	31	35	9,2	43,5	40		29
19399	207,5		23		27			40	38,3	32,5	28,3
19392	209	29,5	24	40,5	28,1	32	15	37,5			

19404	211	31	24,4	42	29	36	12,1	39	38,8	33	30,2
19408	219	32	23					44	40,4	31,5	29
19407	220	29	24	42	27	34,5	11,5	36,5			
19425	203	29,8	22,1	41,7	27,2			38,7	36,5	30,5	27,3
19428	220	31,7	25,5	44,3	29,5	36,8	11,5	43,8	40	33	29,5
19451	216,5	30,5	23,2	44	31,3	36,5		44,3	42,9	32,1	
19423	211	30,5		45	27,5	36,8	12,8	39	38	33	27,8
19448	204	30,5		45	30	36	11,5	40,3	39		
19445	218,8	28,7	23,5	42	21,8			36	35	28,6	26
19443	220,5	30,6	24,4		27,7		11,8	38,9	38,6	32	28,3
19442	226,5	28	24		27		12,5	37			
19453	196,7	28,3	21					37	35,3	32,5	
19435	201,7	28,8	22,7	40,8	28,3	34,5	12	39,6	38,4	31	27,8
19434	220	31,8	24,5					41,4		33,2	
19489	217,4	33,8	26,5	47,7	30,3	40	14,8			34,4	
19480	222	32,4	25,3	44	28,3	36,4	11,5	41	40,4	33,2	29,5
19456	210,7	30	22,7	42,2	27,8	36	11		37,1	29,8	26,5
19459	205	30,5	24		28,8		11,2	39,5	37,4	31,4	27,4
19464	219	30,8	22,8	44,6	30	37	11,4	39,4	39,2	33,5	31,5
19526	219,5	31	22,8	47	30,4	39,8	13,4	42,7	41,2		
19528	208,7	35	24,4	43,2				42,4	40,8	31	29,4
19529	222	30	23,8	40	26,3	33,4	12,3	38	38	32,5	27,4
19547	205,2	30,3	20	8,3	27,6	33,3	15	40	39,5	30,5	26,8
19527	204,2	32,6	22,8	43	26,2	38	12,9	40,5	40,5		26,8
19541	215,5	32,3	22,5	44,1	28	36,3	12	41,2	39,3	33,1	29,6
19543	217	33,7	26,5					40,8	39,7	32,4	30,4
19498	221,2	33,6	25		27,2			40,2	35	31,5	29,2
19504	209	32,3	24,5					40,7			28,2
19517	241,3	29,8	26	45,8	29,4	37,4	12,4	37	38,4	30,5	27,6
19513	207	29,2	22,4	41,5	22,3	35,5	11,5	39		29,3	
19513	210	32,3	23,9					39,5			
19499	211,8	30	23,8	43	28,2	35,4	12	42	40,2	31,3	27,9
19525	210,7	31	24,9	42,6	27,5	34,7	12,8	37,3	36,3		28
19455	222,3	28	22,7	41,5	28,3	34,3	10,8	39,3	38,3	32,1	27,5
17132	221,2	28,2	21,6	44,8	27,4	37,4	13,3	38,2	38	32,6	29,2

Πίνακας-2. *H. brachypus*, Hadjidimono, Βουλγαρία. Μετρήσεις του MCIII (για τις μετρήσεις βλέπε σχ. 6)

	1	3	4	5	6	7	8	16	15	10	11	12	14
I 2D	220	26,5	20,9	39,5	24	31,5	11	0	3	34,1	32	28,3	26
3D	221,7	26		39,7	26,1	34,5	11,5	0	5	34,5	33,5	29,5	27
8D	220			36	24	30,5	10,5			32,3	30,5	26	23,5
16D	213	25,7	21,5	33,5	24	29,5	9	0	4,8	31	30	23,5	21

18=38D	218	26,5	22	35,5	24,3	31,5	9,5	0		32	31,5	26,8	24,5
21D	202	26,2	22	38	25,5	32	10,5		6,2	36,5	34,5	27	25,3
24D	217	25	22	36,5	26,2	32	11	3		32	32,5	27,3	24,5
25D	218	27	23	35,5	25,5	30	10	0		34,5	32	26,2	23
27D	221	25,1	21,5	35,5	23,5	30	9,1		4,1	33,3	32,7	26,4	23,6
B4G	216	24	21,5	36,3	24,3	29,5	11	3,2	5	31,5	31,2	26	23,5
6G	227	24	22	36,5	26,9	32,1	9	4,5	5,5	33,1	32,6	26,9	25,1
7?G	221	26	23,5	36,5	24,5	29,5	10,5			34	31,5	26	24,4
21G	213	26	21	37,3	25	31,5	9,1		5,5		32	26,5	23,5
23G	202	26,5	21	35	24,5	31,5	9,5			32	31	25	22,5
24G	204,5	24,5	20,5	35,5	24	32	9				32	27,1	23
PIK37	213	25	21,3	35,6	25,2	32,8	9		110	32	31,8	26,4	24
PIK51	218	26,5	20,8	37,2	27,8	31,4	9,7	4	104	33,5	32,8	27,7	24,6
PIK53	219	23,7	20,5	37,2	25	30,4	10,8		112	31	30,5	25,1	24
PIK57	214	23,7	20,5	37,1	23,7	31,8	9,6	5,3	103	33,2	31,7	27,5	25
PIK61	206	25,3	20,4	34	21	28,5	8,5	4,2	108	31,2	30,2	29,7	23,6
PIK62	210	26		35	25	31,5	8		107	33	31	24	
PIK150	214	23,8	20	32,8					110	30,4	29,5	25,5	22,8
PIK141	214	23	21,2	39		32,5	9,3		105	34	32,5	27	25,2
PIK154	234	24	22,3	39	26,4	32,5	10,2	4,5	108	34,6	32,5	29,2	27
PIK154a	234	23	21,6	35,7	24	31,6	7,8	5	102	31,5	30,8	26,8	24,5
M11237a	217	28,5	21	40,2	26,4	35	10,4			36,8	34	28,5	
M47902	207	28,7	23,8	40	30	34	9,7			38	35,1	28,9	
M11241a	217,5	29	22,4	38						37,8	35,3	28,5	
M11241b	208	25,3	19,5	27,7	20,7	33,5	8,4			33,8	32	24,6	
M11238	220	25	21,5	34,5	21,8	29,7	9,5			33	31,8	27,4	
M11241f	222	25,5	22,1	37,2	25,1	30,2	9,3			33,9	31,5	26,5	
M11241g	219	24,5	21,4	37,3	23,4	30	9,1			34,5	33	26	
M47850	204	25,3	22,4	36,5	23,1	31,5	10			38,1			
M47845	217	28,7	22	40	26,3	33,8	10,1			37	34,7	29	
M4841	228	24,7	22,2	36,5	22,7	31	9			33,6	33	27	
M47882	212	26,5	22,4		25,2		9,7			33	31,6	25	
M47881	220	25,4	20,9	33	21,6	29,3	8,5			33,8	32,5	26,1	
M11238	220	25,3	21,6	34,7	25	31	9			34	31,9	26,2	
M47871	205	26	22,4	36	25	31,1	9,5			33	33,5	28,6	
W1	218	23,4	21	36,3	26,5	30,5	10,4	3	28				
W2	228	25,5	20,8		23,4	30,2		4,5	34,2	31	27,2	22,2	
W3	216	25,6	21,7	35,5	24,6	30,8	10,6		32,3	32	25	22	
W4	213	24,2	19,8		24,3				30	31	26	22,2	
W5				35	24,2	28,6	12						

Πίνακας-3. *H. mediterraneum*, Πικέρμι, Ελλάδα. Μετρήσεις του MCIII (για τις μετρήσεις βλέπε σχ. 6)

	1	3	4	5	6	7	8	16	15	10	11	12	14
C1=15D	202,5	32,5	22,5	45,5	29	35,5	13,5			39,5		32	30
4=12D	212	31	24,1	42,5	28,5	36,5	12,5			40,5	38,6	30,5	27
5j=13D	212,5	29,5	23,6	42,5						38,5	37,5	30,5	29
6=24D			31	24						40,5	39	29,5	
7=34D	214	31,5	26,5	43,5	29,5	37	11		8	41	39,5	32,4	28,5
9=25D	210	33	24,5	40	27,5	33	12,2	6,8	6,5	40,5		30	27
10=37D	222	31,5	24	42,5	30,5	38,5	13		6	39	36	31,5	28
11=35D	213,5	29,5	24,5	42	30	35	12		7,3	41,5	37	31	27,5
12=29D	218,5	31,5	26,5	42,5	31,5	37	11		5,2	42	40	29,4	26
?=28D	219	29	23	43,5	30	38	10,5	6	5,3	43,5		31,5	28
?=36D	215	30	23	47		34	11			37,5	36,5	32	28,5
?=14D				41		36,5	13,5		6	39,5	37,5	29,5	26,5
15=18D	211	31,5	24	46,5		39	14,5		6,5	39,5	38	28,5	26
19=1?D	206,5	32,5	23,5	42	28,5	35	12		7	39	38,5	28,5	27
20D	210	30	23	40	27,5	34	10,5		4,1	39,2	36	28,7	27,5
22D	209,5	29	23	38,5	29				5,5	38	36,5	29,5	25,5
23D	224	32,2	25	41,5	28,5	34,5	13	4	6	40	38,5	30,7	28
26D	211,5	30	23,1	40,1	28,1	34,5	10,8	5,5	4,2	39,1		29	26,5
28D	216	30,5	25		29		12,3			100,7	39,1	33	29,2
B1=21G	220,5	30	26	43	29,2	35,5	11,5	6,5	5,5	39,5	38,1	30,5	27,5
2=3074G	213	31,5	26,5	44,1	29,8	37	11,5	6,5	8	40,7	39,5	32,2	29,5
3=1974G	230	33,2	25,1	44	29,6	37	11	7	6,8	44,3	41,6	33,2	29
5?G	216	28,7	25,8	40,4	31	34	11,5		6,5	39,1	36,7	29,9	27,3
8G	214	28,5	24,5	41					6,5	39	36,5	29,9	26,8
9G	222,5	34	25	42	32	37,5	10	8		41,2	38,5	32,5	27
10G	205	29,1	22,5	39,5		33,5	11,5		5,5	39	36,5	28	25,3
11G	215	29,8	23,1	43	27,5	35,5	11,5	0	6	39	37,5	30,5	27
12G	198,5	28,5	20,5	37,5	27,5	30,5	11,2	5,1	6,5			28,5	25,3
13G	222	31,5	23,5						57	41,7	39	38,5	30
14G	216	28,5	23,5	42,5	25,8	34,5	12	6		36,2	36	32	28,1
15G	203	33		42	28	33,5	12,5	0		38,5	37,5	30,2	26,5
16G	207	30,1	23,5	41	28,5	34,5	11		6,5	39,5	36,5	31,5	27,5
17G	208	30,5	24,7	41,5	28	33	10	5	6,2	39	36	29,5	27,4
18G	210	31,5	23,5	40	26,5	33,5	12	41	6	39	37	29,5	25,2
20G	203	33	23,5	42,7	30,2	37	13	5,5		41,5	40,6	31,5	27,7
B22G	207,5	32	23,5	38,7	29,5	36,5	10		5,5	39,5	37,5	28,5	26,2
25G	210	32	23,5	40	29	36,5	8		5,5	39,5	38	29	26,2
26G	211	31	24	40,5	30	34	12			39	38,5	29	22,5
27G	205	32	23	42,5		35,5	11,5			41	38	31,5	28,5
28G	212	30,7	22,2	40,6	28,5	36	10			40	38,1	29,5	25,7

PIK-52	205	31,2	24	43,2	28,2	36	11,5		103	41,2	39,7	29,7	26,7
PIK-52a	210	32,7	22,7	43,1	28,5	36,5	11,3		110	42,3	39,4	29,6	28,5
PIK-54	215	35	24,4	40	27,5	32,5	11,3		110	42		29	27,2
PIK-55	207	29,3	21,2	39,4	28	34	9,6		115	38,7	37,5	30	25,3
PIK-59	200	30	22,3	38,2	27,8	32,5	9,3		105	40	37,5	27,6	26,7
PIK-60	226	33	24	43,8	30	37,2	10,6		113	43,2	40,5	32	29
PIK-108	203	29,5	23	37,4		32	10,5		100	38	36,7	26,2	26
PIK-148	213	29	21,4	39	29,4	33	9,7		110	38,3	36,6	29,7	27,3
M11240	203	32	25,2	42	28,8	35	11,7			39,3	40,3	31	
M11237	210	28,3	22,8		27,7					41,1	39,7	28,8	
M11224	204	30,3	23,8	38,5	27,2	32,5	10,1			38,6	36,1	28	
M11239	207	29,2	21,7	41	26,5	35	10,7			39,2	36,2	29,5	
M11240	210	28	23	43		34,2	12,1			38	38,6	30,8	
M11241	220	31,5	25	41,2	27,2	32	11,2			41,4	37,5	30,2	
M11238	213,5	31	25,5	42,6	28,7	34,6	11,5			40,8	37,3	29,8	
M11241c	217,5	24,7	41,8	26	34,7	11,8				40,3	37	31,1	
M11241d	202	30,7	22,6	40	27	34,4	11,2			40,6	36,6	36	
M11241e	197	32,4	22,9	39,5	26	32	11,6			41	36,8	26,7	
M47849	203	30,8	22,5	40,3	27,2	32	11,3			38,7	37	30	
M47848	206	29,6	23,8	27,5	32,2	10,7				40,1	36,4	28	
M47843	215	30,5	23	42	25	33	11,3			41	39,3	28,2	
M47844	216	30,2	22,5	42	29,3	33,8	11,3			38,7	37,7	30	
M47842	219	30	22,5		26		12,4			39,5	39	30,5	
M47846	215	30,6	26,7	43,1	28,5	33	12,7			41,5	38,2	29,1	
M47851	209	27,5	20,6	39,3	27,4	32	10,5			37,5	36,5	28	
M47847	215	32,2	25,2	41,7	29,5	33,5	11,9			40	39,3	30	
M47840	212,5	29,5	22,7	39,6	27	31,6	11,2			39,9	36,8	29,3	
M47852	204,5	29,5	21,9	39	25,5	31,2	11,4			37	36	27,2	
M47839	211	28,4	22,7	38	27,3	30	11,4			39,2	38,6	29	
W1	206	29	22	40,3	27,8	32	11,5	4,6	38,5	36,5	29	23,4	
W2	211	31,7		43	31,2	36	11	5,9	42	38,5	30,6	26,4	
W3									39	37,5	28	23,7	

Πίνακας 4. *H. brachypus*, Πικέρμι, Ελλάδα. Μετρήσεις του MCIII (για τις μετρήσεις βλέπε σχ. 6)

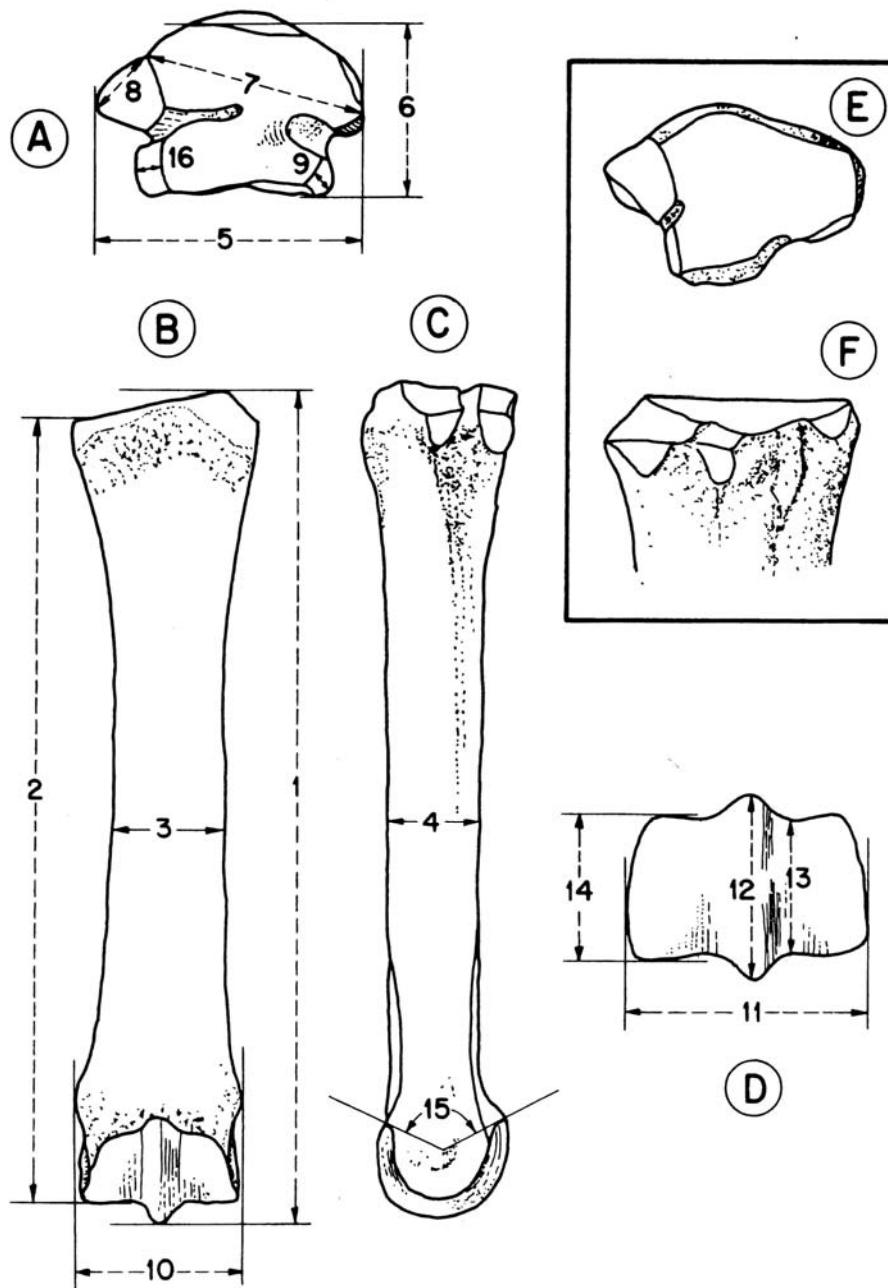
ii. Επεξεργασία των δειγμάτων

Οι περισσότερες πολυπαραγοντικές μέθοδοι ανάλυσης δεδομένων (όπως αυτές που περιγράφηκαν και θα εφαρμοστούν στην συνέχεια) έχουν ως προαπαιτούμενο έναν πίνακα δεδομένων χωρίς κενά κελιά. Έτσι για την περαιτέρω στατιστική ανάλυση είναι απαραίτητη η επεξεργασία των δειγμάτων ώστε τελικά να προκύψουν πλήρεις πίνακες δεδομένων. Αυτό επιτυγχάνεται με την διαγραφή κάποιων στηλών και γραμμών. Δεν υπάρχει ένας συγκεκριμένος τρόπος που πρέπει να εφαρμοστεί σε αυτήν την επεξεργασία ώστε να είναι σωστή, απλώς πρέπει να γίνει με γνώμονα την όσο το δυνατόν μικρότερη έλλειψη πληροφορίας, δηλαδή θα πρέπει να απαλείφονται οι γραμμές και οι στήλες με τα λιγότερα στοιχεία.

Η επεξεργασία αυτή γίνεται έτσι ώστε να προκύψουν τελικά οι ακόλουθοι πίνακες δεδομένων:

1. Απολιθωματοφόρος θέση Hadjidimono (συνδυασμός των ειδών *H. mediterraneum* και *H. brachypus*)
2. Απολιθωματοφόρος θέση Πικέρμι (συνδυασμός των ειδών *H. mediterraneum* και *H. brachypus*)
3. *H. mediterraneum* των θέσεων Hadjidimono και Πικέρμι
4. *H. brachypus* των θέσεων Hadjidimono και Πικέρμι

Οι εφαρμογές των μεθόδων θα πραγματοποιηθούν με το στατιστικό πακέτο PAST (PAleontological STatistics), το οποίο και περιέχει τις περισσότερες από τις στατιστικές μεθόδους χρησιμοποιούνται στην παλαιοντολογία ενώ παράλληλα είναι πολύ εύχρηστο ειδικά για άτομα τα οποία έρχονται πρώτη φορά σε επαφή με ανάλυση δεδομένων.



Σχήμα 6. Η απεικόνιση των μεταβλητών των δειγμάτων και οι περιγραφές τους (από Eisenmann *et al.* 1988) :

- | | |
|--|---|
| 1. Maximal length | magnum |
| 2. Internal length | 9. Diameter of the anterior facet for hamatum |
| 3. Breadth of the diaphysis (in the middle) | 10. Distal maximal supra articular breadth |
| 4. DAP idem at the level of 3 | 11. Distal maximal articular breadth |
| 5. Proximal articular breadth | 12. Distal maximal DAP of the keel |
| 6. Proximal articular DAP | 13. Distal minimal DAP of the lateral condyle |
| 7. Maximal diameter of the articular facet for | 14. Distal maximal DAP of the medial condyle |

III.2 HADJIDIMONO

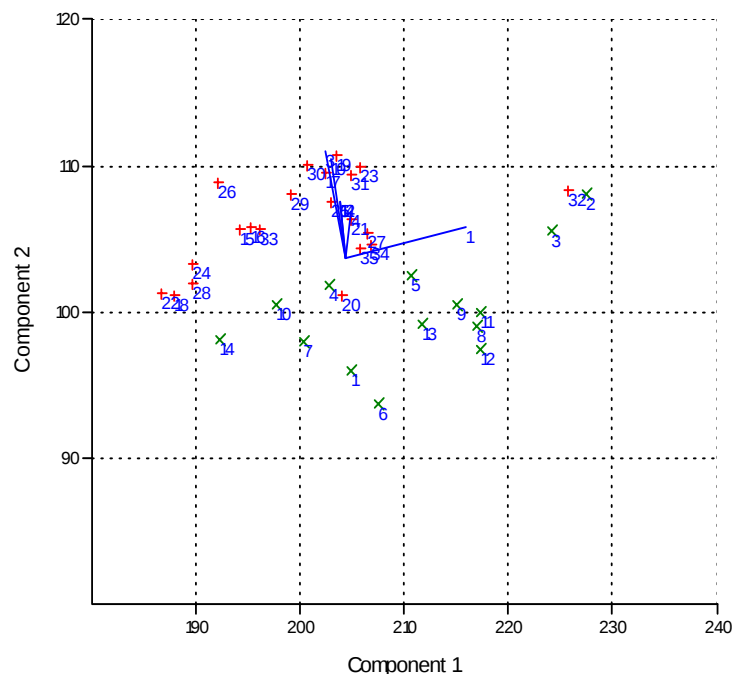
Αφού έχει ολοκληρωθεί η διαδικασία, που αναφέρθηκε παραπάνω, ο τελικός πίνακας δεδομένων στον οποίο θα γίνει η ανάλυση θα περιέχει τις μεταβλητές (στήλες) και τα δείγματα (γραμμές) που είχαν ικανοποιητικό αριθμό μετρήσεων. Για την εξαγωγή συμπερασμάτων και όχι επειδή είναι προαπαιτούμενο των μεθόδων στην κάθε περίπτωση θα δίνεται ένα χρώμα σε κάθε παρατήρηση ανάλογα με το που ανήκει σε κάθε πίνακα δεδομένων. Η μοναδική μέθοδος που απαιτεί για την εφαρμογή της τον εκ των προτέρων διαχωρισμό του δείγματος σε δύο ομάδες είναι η Διακρίνουσα Ανάλυση. Στην πρώτη αυτή εφαρμογή, για τα δείγματα από το Hadjidimono που ανήκουν στο *H. brachypus* θα απεικονίζονται με κόκκινο, ενώ αυτά που ανήκουν στο είδος *H. mediterraneum* με πράσινο.

Αρχικά στο δείγμα θα εφαρμοστεί η Ανάλυση Κύριων Συνιστωσών, που όπως προαναφέρθηκε έχει σκοπό την εύρεση καινούριων μεταβλητών λιγότερων των αρχικών αλλά με βάση αυτές. Οι μεταβλητές θα προκύψουν έτσι ώστε να μην υπάρχει σημαντική μείωση της πληροφορίας και με γνώμονα την καλύτερη απεικόνιση των παρατηρήσεων. Για την μέθοδο χρησιμοποιείται πίνακας διακύμανσης – συνδιακύμανσης από τον οποίο προέκυψαν οι ιδιοτιμές του πίνακα 5 για τις αντίστοιχες κύριες συνιστώσες. Με βάση τις ιδιοτιμές φαίνεται ότι η πρώτη και δεύτερη κύρια συνιστώσα ερμηνεύουν σημαντικό ποσοστό της διακύμανσης του δείγματος (96.33%) με αποτέλεσμα να θεωρούνται αρκετές ώστε να μπορούν να δώσουν μια αξιόπιστη εικόνα για το δείγμα (σχήμα 7). Από το σχήμα 7 φαίνεται ότι τα δυο είδη διαχωρίζονται αρκετά καλά.

	Τιμή	% Διακύμανση
Ιδιοτιμή 1	111.2	81.48
Ιδιοτιμή 2	20.27	14.85
Ιδιοτιμή 3	1.94	1.42
Ιδιοτιμή 4	1.39	1.02

Πίνακας-5. Οι ιδιοτιμές της PCA για τη θέση Hadjidimono.

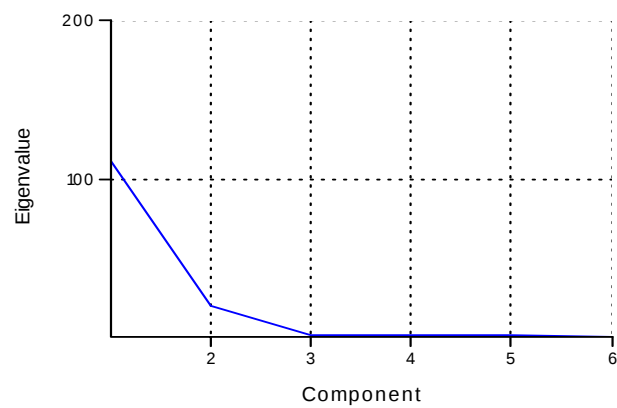
Σημαντικό είναι να βρεθεί ο παράγοντας ή οι παράγοντες που συνεισφέρουν καθοριστικά στον διαχωρισμό των δυο ειδών. Με βάση το γράφημα των μεταβλητών στο κέντρο του σχήματος, που δείχνει την συμπεριφορά τους κατά μήκος των κύριων συνιστωσών, εξάγεται το συμπέρασμα ότι οι μεταβλητές 1 και 3 είναι αυτές που διαφοροποιούνται ανάμεσα στα δυο είδη και κυρίως η 3. Έτσι το είδος *H. brachypus* εμφανίζεται ως πιο κοντό (το πλήθος των παρατηρήσεων συγκεντρώνεται στην αντίθετη πλευρά από όπου αυξάνεται η τιμή της μεταβλητής 1, δηλαδή όπου θα αναμένονται ευρήματα πιο επιμήκη) και πιο εύρωστα (οι παρατηρήσεις βρίσκονται στην κατεύθυνση αύξησης της μεταβλητής 3) σε σχέση με το *H. mediterraneum*. Στην πραγματικότητα στην παρούσα περίπτωση αν και υπάρχει σαφής διαχωρισμός δεν υπάρχει ένας παράγοντας που να τον καθορίζει. Ένας ιδεατός τέτοιος παράγοντας θα ήταν η μεταβλητή της οποίας η διεύθυνση αύξησης θα είναι κάθετη στην ευθεία που χωρίζει τις



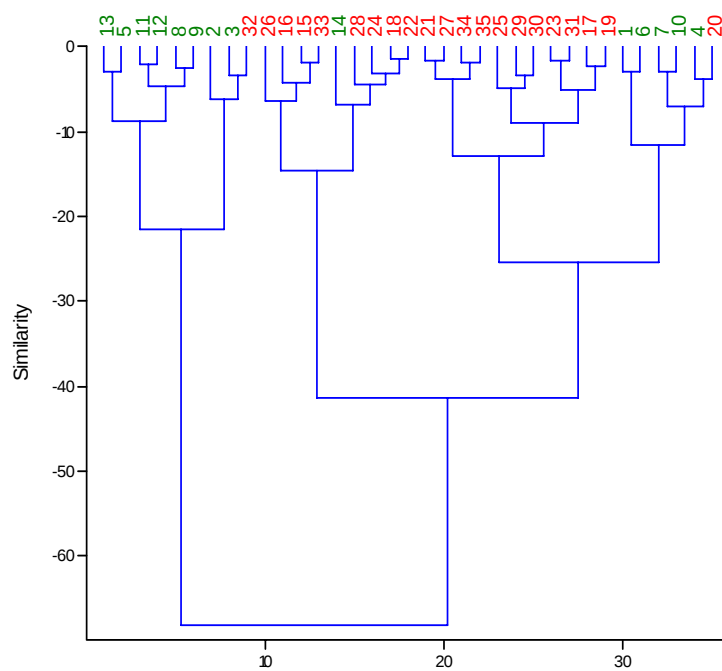
Σχημα-7. Οι πρώτες δύο κύριες συνιστώσες της PCA για το δείγμα Hadjidimono, όπου με κόκκινο απεικονίζεται το *H. brachypus* και με πράσινο το *H. mediterraneum*.

δυο ομάδες και εκτός αυτού εξηγεί σημαντικό ποσοστό της διακύμανσης του δείγματος. Για την εύρεση του απαραίτητου αριθμού κύριων συνιστωσών μπορεί να χρησιμοποιηθεί και το *scree plot* (σχήμα 8) όπου σε αυτήν την μέθοδο λαμβάνονται υπόψη οι κύριες συνιστώσες ως και αυτήν όπου υπάρχει το «σπάσιμο» στο γράφημα, δηλαδή στην προκειμένη περίπτωση στην δεύτερη κύρια συνιστώσα.

Στη συνέχεια της ανάλυσης του δείγματος θα εφαρμοστεί η Ανάλυση Ομάδων σε μια προσπάθεια ομαδοποίησης των παρατηρήσεων του δείγματος με βάση τις ομοιότητές τους. Ο αλγόριθμος που θα χρησιμοποιηθεί είναι του Ward λόγω του ότι φαίνεται να βγάζει καλύτερα αποτελέσματα που μπορούν να ερμηνευθούν ευκολότερα. Η ιεραρχική ταξινόμηση που προκύπτει με βάση αυτή τη μέθοδο φαίνεται στο σχήμα 9. Στο δένδρογραμμα αυτό φαίνεται ένας αρκετά καλός διαχωρισμός των δύο ειδών μοιάζοντας πάρα πολύ με τον διαχωρισμό τους πάνω στο επίπεδο των δυο πρώτων συνιστωσών (σχήμα 7) κατά την εφαρμογή της Ανάλυσης Κύριων Συνιστωσών. Αυτό γίνεται εμφανές με την ύπαρξη των δειγμάτων (1, 4, 6, 7, 10), που έχουν αναγνωρισθεί ποιοτικά ως *H. mediterraneum* να τοποθετούνται στην Ανάλυση Ομάδων καθώς και στην Ανάλυση Κύριων Συνιστωσών στο είδος *H. brachypus*, σχηματίζοντας έτσι μια υποομάδα μέσα σε

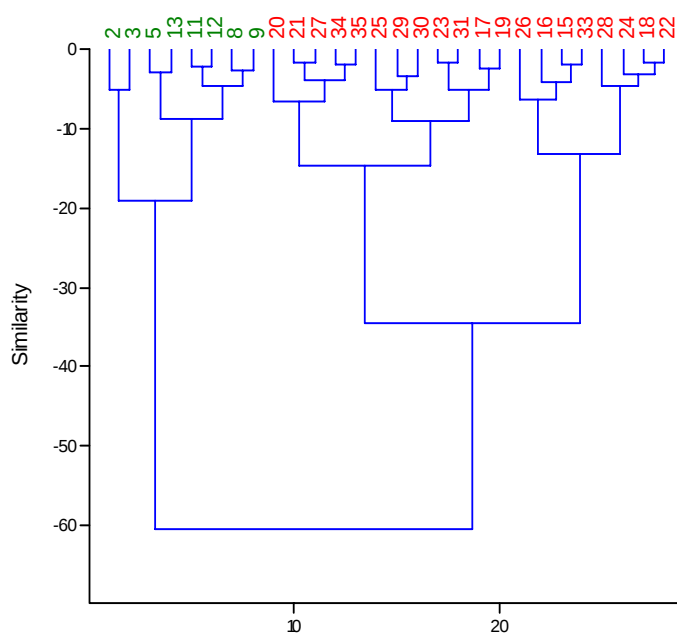


Σχήμα-8. Το *scree plot* της PCA για το δείγμα Hadjidimono.



Σχήμα-9. Το δένδρογραμμα της Ανάλυσης Ομάδων του δείγματος Hadjidimono, όπου με κόκκινο απεικονίζεται το *H. brachypus* και με πράσινο το *H. mediterraneum*.

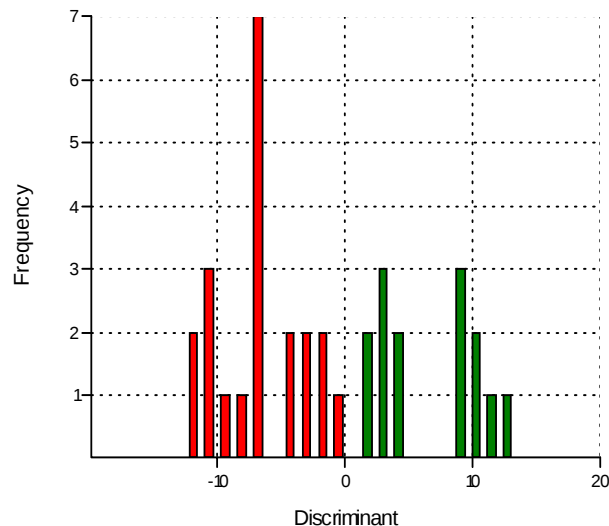
αυτό. Πιθανότατα αυτές οι παρατηρήσεις αν δεν μπορούν να χαρακτηριστούν ως άλλο είδος είναι κάποιες μορφές με ενδιάμεσες διαστάσεις ανάμεσα στα δύο παραπάνω είδη. Ένα άλλο στοιχείο άξιο παρατήρησης είναι το ότι ενώ στο επίπεδο των δυο πρώτων συνιστωσών της Ανάλυσης Κύριων Συνιστωσών το είδος *H. brachypus* είναι αρκετά συγκεντρωμένο (μικρό εύρος τιμών των μεταβλητών μέσα στο είδος) στο δενδρόγραμμα της Ανάλυσης Ομάδων εμφανίζεται ως μια αρκετά ανομοιογενής ομάδα (χαμηλή διακλάδωση της ομάδας) ενώ το αντίθετο συμβαίνει με το είδος *H. mediterraneum*. Αυτή η εικόνα στην Ανάλυση Ομάδων οφείλεται σε δύο παράγοντες: 1) ως είδος εμφανίζει μεγάλο εύρος τιμών στις μεταβλητές που χρησιμοποιήθηκαν και 2) σε κάποιο ποσοστό συμβάλει η ύπαρξη των παρατηρήσεων 1, 4, 6, 7, 10 που σχηματίζουν μια ιδιαίτερη υποομάδα έχοντας τιμές που δεν μπορούν να τα κατατάξουν στατιστικά σε κανένα είδος με σιγουριά. Στο σχήμα 10 φαίνεται το δενδρόγραμμα όπου σε αυτήν την περίπτωση δεν έχουν ληφθεί υπόψη τα παραπάνω δείγματα και φαίνεται μία ομάδα πιο ομοιογενής, όμως σε τέτοιο βαθμό ώστε να θεωρηθεί ότι αυτές οι παρατηρήσεις επηρέαζαν σημαντικά την εικόνα του δενδρογράμματος του σχήματος 9.



Σχήμα-10. Δενδρόγραμμα του Hadjidimono χωρίς τα δείγματα 1,4,6,7,10,14.

Θεωρώντας εκ των προτέρων τα δυο είδη ως δυο διαφορετικά σύνολα θα εφαρμοστεί η Διακρίνουσα Ανάλυση στο συγκεκριμένο δείγμα ώστε να εκφραστεί στατιστικά η βεβαιότητα για τον καλό διαχωρισμό των δυο ειδών. Με απλά λόγια σκοπός της μεθόδου αυτής είναι η εύρεση της ευθείας πάνω στην οποία οι δυο ομάδες

έχουν τον καλύτερο διαχωρισμό και το μικρότερο εύρος τιμών (σχήμα 11). Σε αυτό το σχήμα εκατέρωθεν του 0 βρίσκονται οι δυο ομάδες. Η ισχύς της μεθόδου υπολογίζεται με το ποσοστό των σωστά τοποθετημένων παρατηρήσεων που στην προκειμένη περίπτωση είναι σχεδόν όλες. Ένας δείκτης που δίνει την πιθανότητα οι δύο



Σχήμα-11. Η ευθεία της Διακρίνουσας Συνάρτησης του δείγματος Hadjidimono, όπου με κόκκινο απεικονίζεται το *H. brachypus* και με πράσινο το *H. mediterraneum*.

ομάδες του δείγματος να αποτελούν ουσιαστικά μία είναι η τιμή *Hosteling's T²* που στην παρούσα ανάλυση είναι ίση με $5,25 \cdot 10^{-8}$. Μια τιμή τέτοιας τάξης προσφέρει αρκετή ισχύ στην εφαρμογή της μεθόδου στο παρόν δείγμα με αποτέλεσμα να μπορεί να εξαχθεί το συμπέρασμα ότι είναι διαφορετικά είδη καλά διαχωρισμένα. Ένα βήμα παραπέρα θα είναι η εύρεση της Διακρίνουσας Συνάρτησης ώστε μελλοντικά μη αναγνωρισμένα ευρήματα (με απαραίτητη προϋπόθεση να ανήκουν σε ένα από τα δύο είδη) με γνωστές τις χρησιμοποιημένες στην μέθοδο μεταβλητές να μπορούν να τοποθετηθούν σε κάποιο από τα δύο σύνολα, δηλαδή σε κάποιο είδος. Η Διακρίνουσα Συνάρτηση που προκύπτει είναι η εξής:

$$R=0.0215 \cdot 1 - 2.3845 \cdot 3 + 0.5186 \cdot 4 + 0.3148 \cdot 10 - 2.0798 \cdot 12 + 1.4751 \cdot 14$$

Έτσι για ένα καινούριο εύρημα, το οποίο όπως προαναφέρθηκε είναι γνωστό ότι ανήκει σε ένα από τα δύο είδη, μετριοούνται οι τιμές των συγκεκριμένων μεταβλητών και λύνεται η παραπάνω συνάρτηση. Όπως φαίνεται από το παραπάνω σχήμα 11 για μια τιμή *R*

μεγαλύτερη του 0 το απολίθωμα ανήκει στο είδος *H. mediterraneum*, ενώ για τιμή μικρότερη του στο *H. brachypus*. Αυτή η κατάταξη είναι τόσο βέβαιη όσο έχει υπολογιστεί η ισχύς της μεθόδου.

Σε τέτοιου είδους δείγματα, δηλαδή όταν ο αριθμός των παρατηρήσεων είναι μικρός, τα συμπεράσματα που βγαίνουν πρέπει να υιοθετούνται με πολλή σκεπτικότητα λόγω του γεγονότος ότι στατιστικά η ισχύς ενός συμπεράσματος αυξάνεται με την αύξηση του πλήθους του δείγματος.

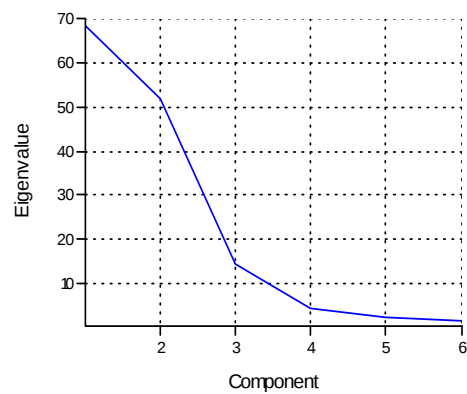
III.3 ΠΙΚΕΡΜΙ

Στην συνέχεια θα αναλυθεί το δείγμα από το Πικέρμι ξεκινώντας με την Ανάλυση Κύριων Συνιστωσών όπου και σε αυτό το δείγμα με πράσινο χρώμα θα πρόκειται για παρατήρηση που ανήκει στο είδος *H. mediterraneum*, ενώ με κόκκινο για *H. brachypus*. Χρησιμοποιώντας και σε αυτή την περίπτωση πίνακα διακύμανσης – συνδιακύμανσης προκύπτουν οι ιδιοτιμές του πίνακα 6 για κάθε κύριο άξονα. Από τον παρακάτω πίνακα

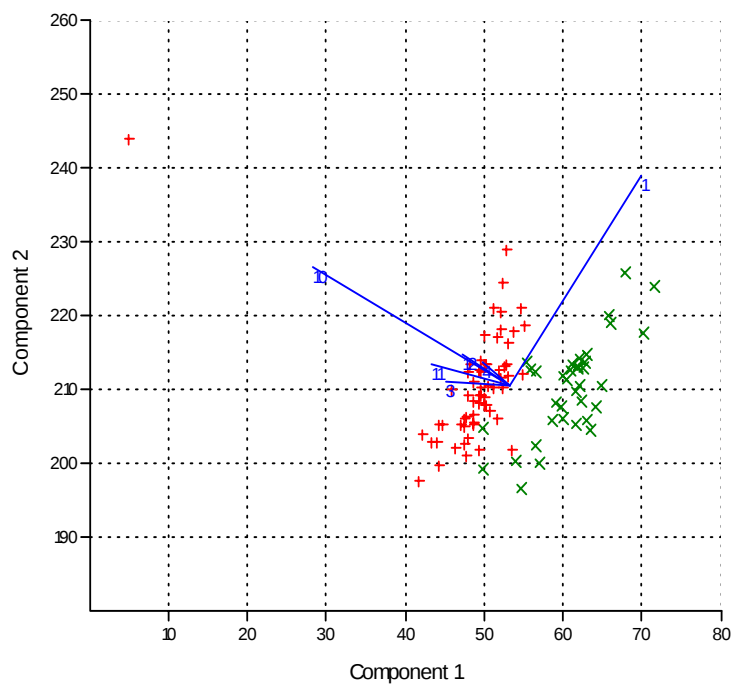
	Τιμή	% διακύμανση
Ιδιοτιμή 1	68,61	47,9
Ιδιοτιμή 2	52,2	36,44
Ιδιοτιμή 3	14,3	9,98
Ιδιοτιμή 4	4,44	3,1

Πίνακας-6. Οι ιδιοτιμές της PCA για τη θέση Πικέρμι.

προκύπτει ότι το 84,34% της ολικής διακύμανσης του δείγματος από το Πικέρμι οφείλεται στους δύο πρώτους κύριους άξονες που προκύπτουν από τη μέθοδο, δίνοντας την δυνατότητα έτσι να μελετηθεί το δείγμα από απεικόνιση των παρατηρήσεων πάνω στο επίπεδο μόνο των δύο πρώτων αξόνων. Είναι χρήσιμο να μελετηθεί και το *scree plot* για την εξαγωγή συμπεράσματος (σχήμα 12). Σε αυτή την περίπτωση είναι εμφανές, με βάση τον εμπειρικό κανόνα που αναλύθηκε παραπάνω, ότι ο απαραίτητος αριθμός κύριων αξόνων που πρέπει να χρησιμοποιηθεί είναι οι δύο πρώτοι. Στο επίπεδο των δυο πρώτων κύριων συνιστωσών (σχήμα 13) γίνεται ένας πάρα πολύ καλός διαχωρισμός των δύο ειδών σχεδόν εκατέρωθεν της διεύθυνσης της μεταβλητής *I*, αυτό σημαίνει ότι η συγκεκριμένη μεταβλητή συμβάλει ελάχιστα στον διαχωρισμό των παρατηρήσεων σε δύο ομάδες. Παρόλη αυτή την αδυναμία του μήκους των οστών να τονίσει ως μεταβλητή την διαφορά ανάμεσα στα δύο είδη εμφανίζεται με μεγάλες τιμές στις παρατηρήσεις του είδους των *H. mediterraneum*, ενώ είναι μειωμένη στην ομάδα των *H. brachypus*. Αντιθέτως, σε αυτό το δείγμα, υπάρχει μία μεταβλητή κατά μήκος της διεύθυνσης της οποίας διαχωρίζονται τα δύο είδη, η μεταβλητή *10*, έτσι ως *H. brachypus* μπορούν να



Σχήμα-12. Το *scree plot* της PCA του δείγματος Πικέρμι.

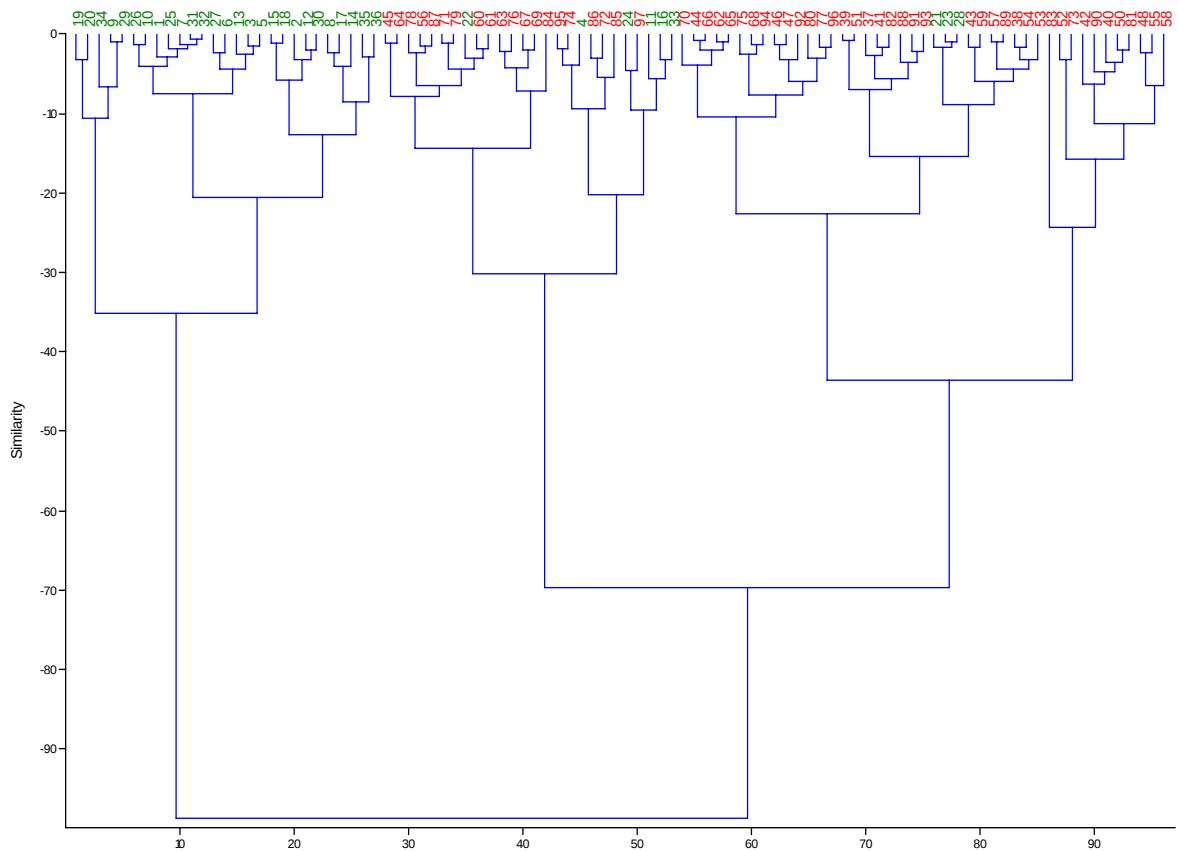


Σχήμα-13. Οι δύο πρώτες κύριες συνιστώσες της PCA του δείγματος Πικέρμι. , όπου με κόκκινο απεικονίζεται το *H. brachypus* και με πράσινο το *H. mediterraneum*.

χαρακτηριστούν τα εύρωστα MCIII του δείγματος ενώ ως *H. mediterraneum* χαρακτηρίζονται τα πιο λεπτά (μικρότερες τιμές πάχους οστών). Μια άλλη παρατήρηση για το επίπεδο των δύο πρώτων κύριων συνιστωσών είναι η απομάκρυνση του δείγματος 28D του *H. brachypus* από τα υπόλοιπα. Σε μια τέτοια περίπτωση πρέπει να βρεθεί η συγκεκριμένη παρατήρηση στον πίνακα δεδομένων (Πίνακας 4) ώστε να ελεγχθεί η ύπαρξη μετρήσεων οι οποίες αποκλίνουν σημαντικά από του υπόλοιπου δείγματος. Στην συγκεκριμένη περίπτωση φαίνεται ότι η τιμή της μεταβλητής **10** της προαναφερθείσας παρατήρησης είναι κατά πολύ μεγαλύτερη σε σχέση με αυτές των υπόλοιπων ευρημάτων. Αυτό το γεγονός με τη σειρά του μπορεί να οφείλεται σε δύο ενδεχόμενα: είτε πρόκειται για λάθος μέτρηση είτε έχει ποιοτικά ταξινομηθεί λάθος το εύρημα, έτσι με βάση τις υπόλοιπες μετρήσεις του ευρήματος αυτού μάλλον ισχύει το πρώτο ενδεχόμενο αφού δεν αποκλίνουν σημαντικά από αυτές των υπόλοιπων παρατηρήσεων. Για την αποφυγή εξαγωγής λάθος συμπερασμάτων δεν θα χρησιμοποιηθεί ξανά στην υπόλοιπη εργασία η παραπάνω παρατήρηση.

Το ίδιο δείγμα θα μελετηθεί με την εφαρμογή της Ανάλυσης Ομάδων, όπου σύμφωνα με την προηγούμενη εξήγηση της μεθόδου διερευνάται η ύπαρξη ομάδων στο δείγμα που ορίζονται από παρατηρήσεις που εμφανίζουν αυξημένη μεταξύ τους ομοιότητα. Στην ανάλυση θα χρησιμοποιηθεί ο αλγόριθμος Ward. Το δένδrogramma που προκύπτει από την εφαρμογή της μεθόδου αφού δεν έχει ληφθεί υπόψη η παρατήρηση 28D φαίνεται στο σχήμα 14. Στο παρακάτω δένδrogramma σχεδόν όλες οι παρατηρήσεις που ανήκουν στο είδος *H. mediterraneum* ορίζουν μια ομάδα, η οποία λόγω του ότι διακλαδίζεται πολύ ψηλά δηλώνει ένα μεγάλο βαθμό ομοιότητας ανάμεσα στα μέλη της. Από την άλλη μεριά η δεύτερη ομάδα που δημιουργείται διακλαδίζεται αρκετά χαμηλά δηλώνοντας έτσι ότι τα μέλη της έχουν τιμές που παρεκκλίνουν αρκετά μεταξύ τους. Η τελευταία αυτή ομάδα αποτελείται από όλες τις παρατηρήσεις που ανήκουν στο είδος *H. brachypus* και ελάχιστες από το *H. mediterraneum*, καθιστώντας έτσι τον διαχωρισμό των ομάδων από την συγκεκριμένη μέθοδο ικανοποιητικά καλό. Με βάση αυτές τις παρατηρήσεις δημιουργείται η εικόνα ενός είδους *H. brachypus* με πολύ «διεσπαρμένες» τιμές ενώ το αντίθετο συμβαίνει όσον αφορά το *H. mediterraneum*, γεγονός το οποίο δεν φαίνεται καλά στην εφαρμογή της Ανάλυσης Κύριων Συνιστωσών για το επίπεδο των δυο πρώτων κύριων αξόνων.

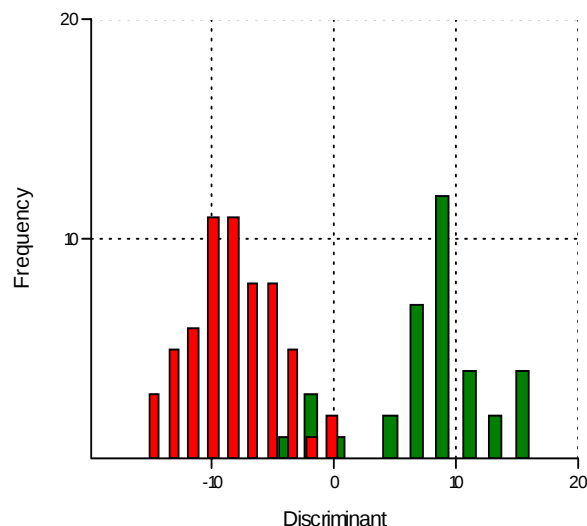
Τελευταία μέθοδος που θα εφαρμοστεί για αυτό το δείγμα είναι η Διακρίνουσα Ανάλυση, ώστε να βρεθεί το επίπεδο βεβαιότητας με το οποίο υπάρχει διαχωρισμός ανάμεσα στα δύο είδη. Ο διαχωρισμός (σχήμα 15) φαίνεται να είναι αρκετά καλός με



Σχήμα-14. Δενδρόγραμμα του δείγματος Πικέρμι, όπου με κόκκινο απεικονίζεται το *H. brachypus* και με πράσινο το *H. mediterraneum*.

μόνη εξαίρεση κάποια μέλη του είδους *H. mediterraneum*, τα οποία βρίσκονται κατά μήκος της ευθείας της διακρίνουσας συνάρτησης σε τιμές μικρότερες του 0, δηλαδή στην περιοχή του είδους *H. brachypus*, πράγμα το οποίο ήταν αναμενόμενο από την Ανάλυση Ομάδων του ίδιου δείγματος. Η πιθανότητα οι δυο αυτές ομάδες να είναι στην πραγματικότητα μία ισούται με $2,04 \cdot 10^{-29}$, αποκλείοντας έτσι αυτό το ενδεχόμενο. Τέλος η διακρίνουσα συνάρτηση είναι:

$$P = 0,1992 \cdot 1 - 0,8653 \cdot 3 - 0,3554 \cdot 4 - 1,3542 \cdot 10 - 0,3925 \cdot 11 - 0,0356 \cdot 12$$



Σχήμα-15. Η Διακρίνουσα Ανάλυση του δείγματος Πικέρμι, όπου με κόκκινο απεικονίζεται το *H. brachypus* και με πράσινο το *H. mediterraneum*.

III.4 *H. MEDITERRANEUM*

Σε αυτό το δείγμα έχουν τοποθετηθεί όλα τα ευρήματα του είδους *H. mediterraneum* που έχουν βρεθεί και στις δύο απολιθωματοφόρες θέσεις (Hadjidimono και Πικέρμι). Με αυτό τον τρόπο μπορούν να βρεθούν τυχόν διαφορές μέσα στο ίδιο είδος, οι οποίες να οφείλονται σε τρεις παράγοντες: α) τη διαφορετική γεωγραφία β) τα διαφορετικά περιβάλλοντα και γ) τη διαφορετική ηλικία. Σε αυτό αλλά και στο επόμενο δείγμα οι ομάδες πλέον που μελετώνται για την διαφορετικότητά τους είναι οι τοποθεσίες και όχι τα είδη. Έτσι παρατηρήσεις οι οποίες προέρχονται από τη θέση Hadjidimono θα απεικονίζονται με μπλε χρώμα, ενώ αυτές που προέρχονται από το Πικέρμι με μωβ.

Πριν την ανάλυση του δείγματος πρέπει να δοθεί προσοχή στο γεγονός ότι το πλήθος των ευρημάτων που ανήκουν στο είδος *H. mediterraneum* και προέρχονται από τη θέση Hadjidimono είναι μικρό, άρα τα συμπεράσματα θα είναι απαραίτητο σε αυτή τη περίπτωση να συγκριθούν με ποιοτικές αναλύσεις του δείγματος.

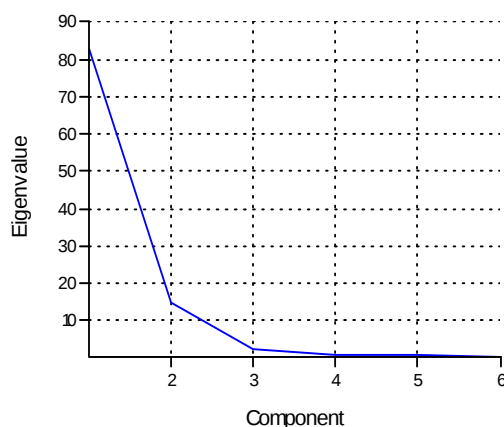
Η μελέτη του δείγματος και σε αυτήν την περίπτωση ξεκινάει με Ανάλυση Κύριων Συνιστωσών, με τις τιμές του πίνακα 7. Οι ιδιοτιμές των δύο πρώτων κύριων

συνιστωσών είναι αρκετές ώστε να δώσουν ένα μεγάλο ποσοστό διακύμανσης (95,28%) για αυτές επί του ολικού δείγματος και

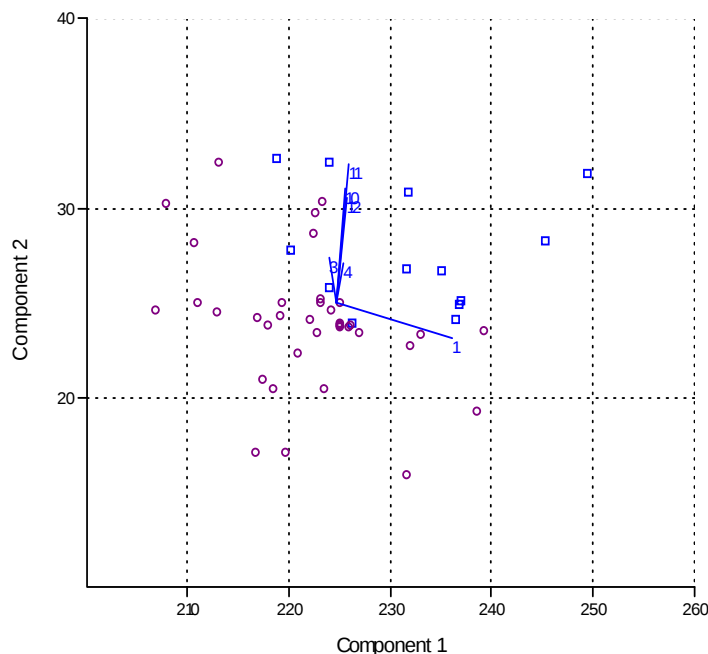
	Τιμή	% διακύμανση
Ιδιοτιμή 1	83,01	80,92
Ιδιοτιμή 2	14,73	14,36
Ιδιοτιμή 3	2,51	2,44
Ιδιοτιμή 4	0,99	0,97

Πίνακας-7. Οι ιδιοτιμές της PCA για το είδος *H. mediterraneum*.

αρκούν για την απεικόνιση του. Το ίδιο συμπέρασμα εξάγεται και από το *scree plot* (σχήμα 16). Στη συνέχεια (σχήμα 17) φαίνονται οι παρατηρήσεις τοποθετημένες πάνω στο επίπεδο που ορίζεται από τις δύο πρώτες συνιστώσες, όπου υπάρχει σαφής διαχωρισμός των δύο τοποθεσιών. Θα μπορούσε να ειπωθεί ότι τα ευρήματα του *H. mediterraneum* που προέρχονται από την απολιθωματοφόρο θέση του Hadjidimono καταλαμβάνουν το πάνω και δεξιό τμήμα του επιπέδου, ενώ αυτά από το Πικέρμι το κάτω αριστερό. Η παρατήρηση αυτή έχει ως συνέπεια να θεωρηθεί ότι η θέση του Hadjidimono έχει μορφές *H. mediterraneum* οι οποίες σε γενικές γραμμές χαρακτηρίζονται από πιο πλατιά μεταπόδια σε σχέση με αυτές του Πικερμίου. Αυτή η διαφοροποίηση θα μπορούσε να αποδοθεί σε δύο ενδεχόμενα: είτε υπάρχει χρονολογική διαφορά των δυο απολιθωματοφόρων θέσεων, είτε αυτή η διαφοροποίηση οφείλεται σε διαφορετικά περιβάλλοντα. Άλλο ένα σχόλιο είναι το πολύ μικρό εύρος τιμών που εμφανίζει το *H. mediterraneum* στο Hadjidimono σε σχέση με αυτό του Πικερμίου.



Σχήμα-16. Το *scree plot* της PCA του είδους *H. mediterraneum*.

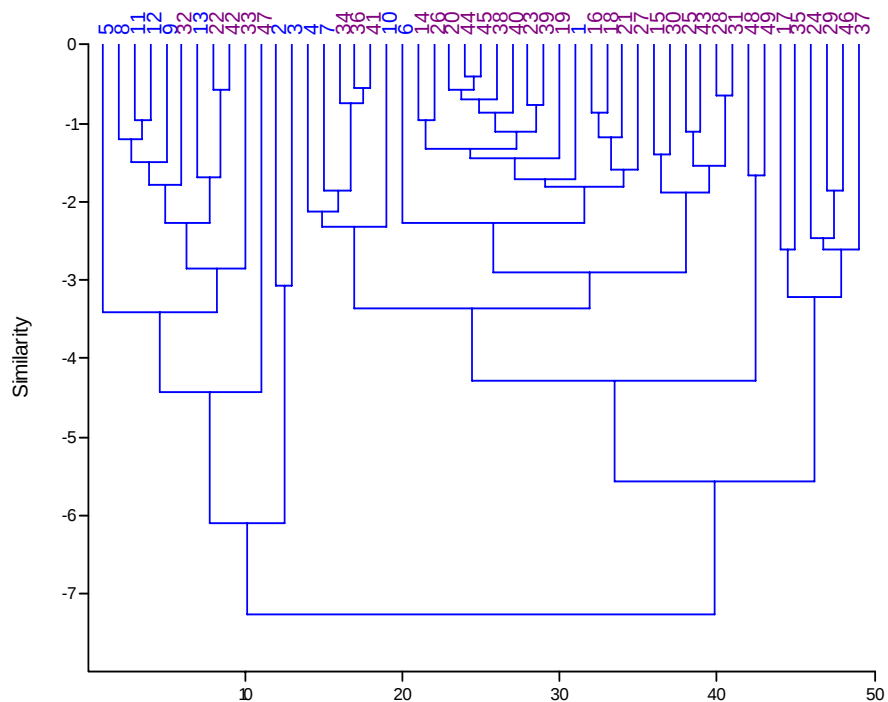


Σχήμα-17. Οι δύο πρώτες κύριες συνιστώσες της PCA του είδους *H. mediterraneum*, όπου με μωβ απεικονίζονται τα δείγματα που προέρχονται από το Πικέρμι ενώ με μπλέ αυτά που προέρχονται από το Hadjidimono.

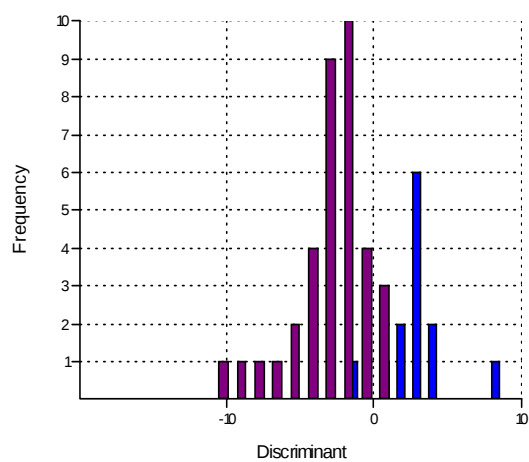
Το δένδρογραμμα (σχήμα 18) από την Ανάλυση ομάδων, που θα εφαρμοστεί με τον αλγόριθμο *paired group*, δεν δίνει τόσο ικανοποιητικό διαχωρισμό όσο η Ανάλυση Κύριων Συνιστωσών. Το δένδρογραμμα οδηγεί σε δύο ομάδες των οποίων τα μέλη πρώτον είναι πολύ ανομοιογενή (χαμηλή διακλάδωση της ομάδας άρα μικρή ομοιότητα των υποσυνόλων της) και δεύτερον είναι ευρήματα που προέρχονται και από τις δύο θέσεις.

Τέλος θα εφαρμοστεί η Διακρίνουσα Ανάλυση για να γίνει σαφές αν τελικά μέσα σε αυτό το δείγμα μπορούν να διαχωριστούν τα ευρήματα των δυο απολιθωματοφόρων θέσεων. Υπάρχει διαχωρισμός των δύο ομάδων (σχήμα 19) με σφάλμα τύπου I $3,41 \cdot 10^{-6}$, δηλαδή πρόκειται για ένα αρκετά ασφαλές συμπέρασμα, ενώ η διακρίνουσα συνάρτηση είναι η εξής:

$$R = 0,096 \cdot I - 0,843 \cdot 3 - 0,159 \cdot 4 + 0,107 \cdot 10 + 1,049 \cdot 11 + 0,14 \cdot 12$$



Σχήμα-18. Το δενδρόγραμμα για το είδος *H. mediterraneum*, όπου με μωβ απεικονίζονται τα δείγματα που προέρχονται από το Πικέρμι ενώ με μπλέ αυτά που προέρχονται από το Hadjidimono.



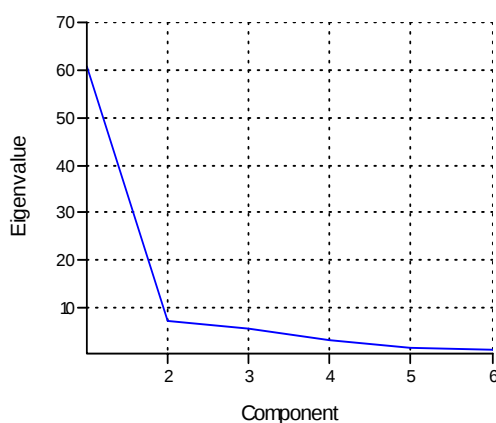
Σχήμα-19. Διακρίνουσα ανάλυση του *H. mediterraneum*, όπου με μωβ απεικονίζονται τα δείγματα που προέρχονται από το Πικέρμι ενώ με μπλέ αυτά που προέρχονται από το Hadjidimono.

III.5 *H. BRACHYPUS*

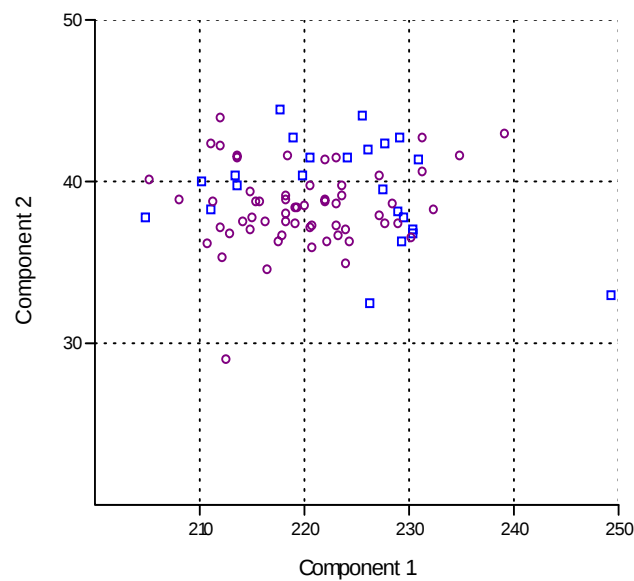
Σε αυτό το δείγμα, όπως και στο προηγούμενο, θα μελετηθεί ο πιθανός διαχωρισμός ανάμεσα στα ευρήματα του *H. brachypus* ανάμεσα στις δύο απολιθωματοφόρες θέσεις. Ο πίνακας ιδιοτιμών της Ανάλυσης Κύριων Συνιστωσών (πίνακας 8) δίνει για τις δύο πρώτες κύριες συνιστώσες ένα ποσοστό 86,4% επί της ολικής διακύμανσης του δείγματος και κρίνεται επαρκής για εξαγωγή συμπερασμάτων σε ολόκληρο το δείγμα, γεγονός το οποίο επιβεβαιώνεται και από το *scree plot* (σχήμα 20) της ανάλυσης. Με βάση το επίπεδο των δύο πρώτων αξόνων (σχήμα 21) γίνεται εμφανές ότι δεν υπάρχει διαχωρισμός μεταξύ των δειγμάτων των δύο θέσεων. Αυτή η ταύτιση των μορφών ανάμεσα στις δύο περιοχές μπορεί να οφείλεται είτε στην ίδια ηλικία των ευρημάτων, είτε στην παρόμοια οικολογία των δύο θέσεων.

	Τιμή	%διακύμανση
Ιδιοτιμή 1	60,92	77,2
Ιδιοτιμή 2	7,26	9,2
Ιδιοτιμή 3	5,43	6,88
Ιδιοτιμή 4	2,89	3,66

Πίνακας-8. Ιδιοτιμές της PCA του είδους *H. brachypus*.

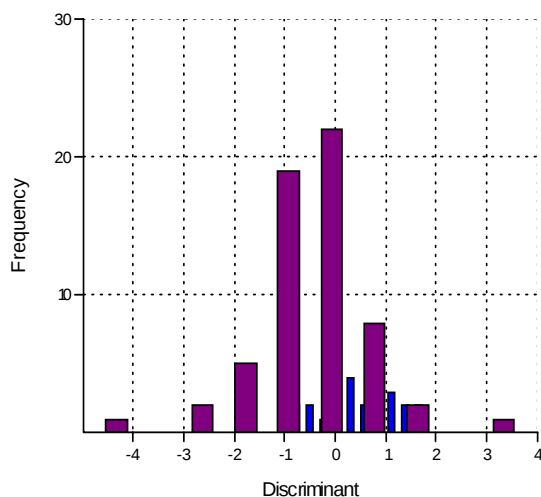


Σχήμα-20. Το *scree plot* της PCA του είδους *H. brachypus*.



Σχήμα-21. Οι δύο πρώτες κύριες συνιστώσες της PCA του είδους *H. brachypus*, όπου με μωβ απεικονίζονται τα δείγματα που προέρχονται από το Πικέρμι ενώ με μπλέ αυτά που προέρχονται από το Hadjidimono.

Την επιβεβαίωση της προαναφερθείσας ταύτισης δίνει και η Διακρίνουσα Ανάλυση (σχήμα 22) στην οποία το σφάλμα τύπου I είναι 0,08 μην επιτρέποντας τον χωρισμό των παρατηρήσεων του δείγματος σε δύο ομάδες. Σε αυτή την περίπτωση δεν είναι χρήσιμη η διακρίνουσα συνάρτηση εφόσον δεν διακρίνονται ομάδες.



Σχήμα 22. Διακρίνουσα Ανάλυση του είδους *H. brachypus*, όπου με μωβ απεικονίζονται τα δείγματα που προέρχονται από το Πικέρμι ενώ με μπλέ αυτά που προέρχονται από το Hadjdimono.

IV. Σύνοψη – Συμπεράσματα

Με βάση τα παραπάνω γίνεται σαφές ότι οι πολυπαγοντικές μέθοδοι χρησιμοποιούνται τόσο για να επιβεβαιώσουν τις ποιοτικές αναλύσεις όσο και να φανερώσουν κάποιες πληροφορίες οι οποίες είναι δύσκολο να γίνουν αντιληπτές μόνο με το μάτι. Αυτή η ικανότητα των μεθόδων οφείλεται στην ποσοτικοποίηση των παραμέτρων με αποτέλεσμα κάποιες αφηρημένες έννοιες (όπως βραχύ, μακρύ κ.α.) ως τώρα να μπορούν να εκφραστούν πλέον ως ένα εύρος τιμών. Βέβαια από την άλλη μεριά δεν είναι δυνατή η πλήρης ποσοτικοποίηση της επιστήμης της Παλαιοντολογίας λόγω του ότι οι μορφές των οργανισμών είναι τόσο πολύπλοκες ώστε να είναι αδύνατο να εκφραστούν αποκλειστικά και μόνο με κάποιο πλήθος τιμών. Προς το παρόν αυτό που μπορεί να κάνει η ποσοτική ανάλυση είναι να συνοδεύσει και να ενισχύσει μια ποιοτική.

V. ΠΑΡΑΡΤΗΜΑ

A.1 Ιδιοτιμές (Eigenvalues) - Ιδιοδιανύσματα (Eigenvectors)

Έχουμε την παρακάτω εξίσωση πινάκων:

$$[A] \cdot [X] = \lambda [X] \quad (\text{A.2.1})$$

όπου $[A]$ είναι ένας πίνακας διακύμανσης – συνδιακύμανσης ή συσχέτισης. Σκοπός μας είναι να βρούμε για ποιες τιμές του λ ισχύει η παραπάνω σχέση η οποία μπορεί να γραφτεί και ως:

$$([A] - \lambda [I]) \cdot [X] = [0] \quad (\text{A.2.2})$$

υπάρχουν δύο περιπτώσεις στις οποίες αληθεύει η παραπάνω εξίσωση, η πρώτη είναι να έχουμε $[X]=[0]$ ενώ η δεύτερη είναι να ισχύει:

$$|A - \lambda I| = 0 \quad (\text{A.2.3})$$

δηλαδή να μηδενίζεται η ορίζουσα του πίνακα $[A - \lambda I]$. Τα λ τα οποία ικανοποιούν τη παραπάνω σχέση αποτελούν τις **ιδιοτιμές**, το πλήθος των οποίων ισούται με τις διαστάσεις του πίνακα. Οι ρίζες αυτές υπάρχει πιθανότητα να είναι μιγαδικοί αριθμοί στην περίπτωση μόνο που ο πίνακας A δεν είναι τετραγωνικός. Στις περιπτώσεις εφαρμογής στην Παλαιοντολογία και τη Βιολογία ο πίνακας όντας συσχέτισης ή διακύμανσης – συνδιακύμανσης είναι πάντοτε τετραγωνικός.

Παίρνοντας τις ιδιοτιμές και λύνοντας τη σχέση A.2.2 βρίσκουμε το διάνυσμα X το οποίο είναι ένα **ιδιοδιάνυσμα**. Είναι εύλογο ότι ο αριθμός των ιδιοτιμών και των ιδιοδιανυσμάτων θα είναι ίδιος.

A.2 Πίνακας διακύμανσης – συνδιακύμανσης (variance – covariance matrix) και συσχέτισης (correlation matrix)

Έστω πίνακας παρατηρήσεων $X(i,j)$, αν N είναι το πλήθος των παρατηρήσεων τότε τα στοιχεία ενός **πίνακα διακύμανσης – συνδιακύμανσης** βρίσκονται από τον τύπο:

$$s_{ij} = \frac{\sum_{n=1}^N (x_{ni} - \bar{x}_i)(x_{nj} - \bar{x}_j)}{N} \quad (\text{A.1.1})$$

ενώ τα στοιχεία ενός **πίνακα συσχέτισης** είναι:

$$r_{ij} = \frac{S_{ij}}{S_i \cdot S_j} \quad (\text{A.1.2})$$

B.1 Πίνακας συγχωνευμένων διακυμάνσεων και συνδιακυμάνσεων.

Για την εύρεση του πίνακα pooled διακυμάνσεων και συνδιακυμάνσεων πρέπει να υπολογιστεί αρχικά ο πίνακας αθροισμάτων τετραγώνων και εξωτερικών γινομένων όλων των μεταβλητών της ομάδας A και αναλόγως της ομάδας B , έτσι για την ομάδα A θα ισχύει:

$$SPA_{jk} = \sum_{i=1}^{n_a} (A_{ij} \cdot A_{ik}) - \frac{\sum_{i=1}^{n_a} B_{ij} \sum_{i=1}^{n_a} A_{ik}}{n_a}$$

όπου A_{ij} συμβολίζει την i παρατήρηση για την j μεταβλητή της A ομάδας, A_{ik} είναι η i παρατήρηση της μεταβλητής k στην ίδια ομάδα. Ανάλογα ισχύει και για την ομάδα B . Έτσι ο πίνακας συγχωνευμένων διακυμάνσεων και συνδιακυμάνσεων βρίσκεται ως εξής:

$$[S_p^2] = \frac{[SPA] + [SPB]}{n_a + n_b - 2}$$

B.2 Πίνακας διαφορών μέσων όρων δυο ομάδων

Ο πίνακας των διαφορών των μέσων όρων δύο ομάδων βρίσκεται από τον τύπο:

$$D_j = \frac{\sum_{i=1}^{n_a} A_{ij}}{n_a} - \frac{\sum_{i=1}^{n_b} B_{ij}}{n_b}$$

όπου A_{ij} είναι η i παρατήρηση της j μεταβλητής στην ομάδα A με πλήθος παρατηρήσεων n_a , το αντίστοιχο ισχύει και για την ομάδα B .

Γ.1 Τυποποίηση Δεδομένων

Έστω πίνακας δεδομένων $n \times m$ όπου n οι παρατηρήσεις και m οι μεταβλητές. Για την τυποποίηση των μετρήσεων χρησιμοποιείται ο παρακάτω τύπος:

$$w_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

όπου w_{ij} η τυποποιημένη παρατήρηση της αρχικής x_{ij} που είναι το στοιχείο του πίνακα που βρίσκεται στη σειρά i και τη στήλη j , $\bar{x}_j$ είναι η μέση τιμή του δείγματος για την μεταβλητή j δηλαδή πιο απλά είναι ο μέσος όρος της στήλης j , ενώ s είναι η διακύμανση για την ίδια μεταβλητή. Με αυτό τον τρόπο ένα σύνολο παρατηρήσεων μετατρέπεται σε μονάδες διακύμανσης.

VI. Βιβλιογραφία

- Cox, T. F. & Cox, M. A. A. 2001. Multidimensional Scaling 2nd edition. Chapman & Hall.
- Davis, J. C. 1986. Statistics and Data Analysis in Geology 2nd edition. John Wiley & Sons.
- Eisenmann, V., Alberdi, M. T., De Giuli, C., Staesche, U. 1988. Methodology. In: Woodburne, M. & Sondaar, P. Y. (eds). Studying Fossil Horses. pp. 1-71. E. J. Brill press, Leiden and New York
- Hammer, O. & Harper, D. A. T. Paleontological Data Analysis.
- Harper, D. A. T. (ed) 1999. Numerical Palaeobiology. John Wiley & Sons.
- Reyment, R. A. 1991. Multidimensional Palaeobiology. Pergamon Press.
- Sokal, R. R. & Rohlf, F. J. 1995. Biometry 3rd edition.
- Vlachou, T. & Koufos G. D. 2002. The hipparions (Mammalia, Perissodactyla) from the Turolian locality "Nikiti-2" (NIK), Macedonia, N. Greece. Annales de Paléontologie, **88**: 215-263.
- Βουδούρης, Κ. 2004. Στατιστική Ανάλυση Δεδομένων στην Εφαρμοσμένη και Περιβαλλοντική Γεωλογία. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.
- Καρλής, Δ. 2003. Πολυμεταβλητή Στατιστική Ανάλυση. Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών.
- Κουφός, Γ. Σημειώσεις Παλαιοντολογίας των Σπονδυλωτών. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης