

# Δείκτης Σύγκλισης Ομάδων Δεδομένων ως προς Ομάδα “Στόχο”: Εφαρμογή σε Δεδομένα Ενεργειακών Εισροών-Εκροών σε Αγροοικοσυστήματα

Γ. Μενεξές<sup>1</sup>, Α. Μάρκος<sup>2</sup>, Ν. Φαρμάκης<sup>3</sup>

*1* Εργαστήριο Γεωργίας, Γεωπονική Σχολή, ΑΠΘ  
gmenexes@agro.auth.gr

*2* Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης, Εργαστήριο Μαθηματικών  
& Πληροφορικής, amarkos@eled.duth.gr

*3* Μαθηματικό Τμήμα, Σχολή Θετικών Επιστημών, ΑΠΘ  
farmakis@math.auth.gr

## Περίληψη

Σε πρακτικές εφαρμογές απαιτείται η σύγκριση δύο ή περισσότερων ομάδων “αντικειμένων”, τα οποία χαρακτηρίζονται από κοινό πλήθος ιδιοτήτων ή χαρακτηριστικών (μεταβλητών), ως προς μία ομάδα “στόχο”. Ο όρος “αντικείμενα” αναφέρεται σε άτομα, πράγματα, γεγονότα ή γενικά σε οποιεσδήποτε οντότητες για τις οποίες υπάρχουν καταγραφές μετρήσεων. Οι τιμές των μεταβλητών της ομάδας “στόχου” είτε αποτελούν σημεία αναφοράς για τη σύγκριση των υπολοίπων ομάδων είτε αποτελούν θεωρητικές-ιδανικές τιμές για την ικανοποίηση συγκεκριμένων κριτηρίων, τα οποία καθορίζονται με βάση το θεωρητικό πλαίσιο της εκάστοτε μελέτης. Στην περίπτωση αυτή, είναι χρήσιμη η κατασκευή και χρήση συναρτήσεων-δεικτών, οι οποίες μετρούν το βαθμό προσέγγισης-σύγκλισης (ομοιότητας) ή απομάκρυνσης-απόκλισης (ανομοιότητας) των ομάδων από την ομάδα “στόχο”. Έστω ότι έχουμε  $c$  ομάδες αντικειμένων, οι οποίες περιγράφονται από πίνακες της μορφής «αντικείμενα · μεταβλητές». Οι τιμές των μεταβλητών αντιστοιχούν στις διαθέσιμες μετρήσεις, οι οποίες αφορούν, για κάθε αντικείμενο, είτε σε διαφορετικές ιδιότητες είτε και στις ίδιες αλλά κάτω από διαφορετικές ή καταστάσεις-συνθήκες στο χώρο και στο χρόνο. Το γενικό πρόβλημα στο οποίο επιχειρούμε να δώσουμε λύση είναι το ακόλουθο: Να κατασκευαστεί συνάρτηση-δείκτης που να μετρά το βαθμό σύγκλισης

(ομοιότητας) ή απομάκρυνσης (ανομοιότητας) των  $c-1$  ομάδων από την ομάδα “στόχο”  $t$ , με βάση τις τιμές των  $p$  παραμέτρων-μεταβλητών, αφού προηγουμένως έχει απαλειφθεί από αυτές η όποια επίδραση έχει επιφέρει η ομαδοποίηση των αντικειμένων. Στην παρούσα εργασία προτείνουμε μεθοδολογία επίλυσης του προβλήματος βασισμένη στην «Προκρούστια Ορθογώνια Προβολή ή Περιστροφή - *Procrustes Orthogonal Projection or Rotation*». Η μέθοδος εφαρμόστηκε σε δεδομένα ενεργειακών εισροών-εκροών 5 αγροοικοσυστημάτων.

### 1. Εισαγωγή-Το Γενικό Πρόβλημα

Σε εμπειρικές έρευνες ή πειράματα συχνά απαιτείται η σύγκριση δύο ή περισσότερων ομάδων “αντικειμένων”, τα οποία χαρακτηρίζονται από κοινό πλήθος ιδιοτήτων ή χαρακτηριστικών (παραμέτρων-μεταβλητών), ως προς μία ομάδα “στόχο” (Dijksterhuis, Martens & Martens 2005, Andrade *et al.* 2004). Ο όρος “αντικείμενα” αναφέρεται σε άτομα, ζώα, φυτά, πράγματα, γεγονότα ή, γενικά, σε οποιεσδήποτε οντότητες για τις οποίες υπάρχουν καταγραφές μετρήσεων. Οι τιμές των παραμέτρων της ομάδας “στόχο” είτε αποτελούν σημεία αναφοράς για τη σύγκριση των υπολοίπων ομάδων είτε αποτελούν θεωρητικές-ιδανικές τιμές για την ικανοποίηση συγκεκριμένων αντικειμενικών κριτηρίων (ή προϋποθέσεων), τα οποία καθορίζονται με βάση το θεωρητικό πλαίσιο και το σκοπό της εκάστοτε μελέτης. Στην περίπτωση αυτή είναι χρήσιμη η κατασκευή συναρτήσεων-δεικτών, οι οποίες μετρούν το βαθμό προσέγγισης (σύγκλισης, ομοιότητας) ή απομάκρυνσης (απόκλισης, ανομοιότητας) των ομάδων από την ομάδα “στόχο”. Στις ενότητες που ακολουθούν, αρχικά γίνεται μια σύντομη αναφορά σχετικά με τις επιπτώσεις στην ερμηνεία των αποτελεσμάτων από την πιθανή επίδραση της ομαδοποίησης των αντικειμένων (Ενότητα 2). Στη συνέχεια, διατυπώνεται, εξειδικεύεται και επιλύεται το αντίστοιχο μαθηματικό πρόβλημα με βάση την Προκρούστια απόσταση (Ενότητες 3 και 4). Στο Παράρτημα Α αποδεικνύεται ότι η Προκρούστια απόσταση είναι μετρική. Τέλος, στο Παράρτημα Β παρουσιάζεται ένα παράδειγμα εφαρμογής της προτεινόμενης μεθοδολογίας με δεδομένα ενεργειακών εισροών-εκροών σε συμβατικά και βιολογικά αγροοικοσυστήματα.

### 2. Επίδραση της Ομαδοποίησης των Αντικειμένων

Τα κριτήρια ομαδοποίησης καθορίζονται από χαρακτηριστικά ή ιδιότητες (Kirk 1995, Kutner *et al.* 2005):

- α) Των αντικειμένων. Για παράδειγμα, αν πρόκειται για άτομα, παράγοντες ομαδοποίησης συνήθως αποτελούν το φύλο, η ηλικία, το εισόδημα, η ευφυΐα, το μορφωτικό επίπεδο, η επαγγελματική εμπειρία και οι συνήθειές τους. Στην περίπτωση γεωγραφικών περιοχών, η ομαδοποίηση μπορεί να είναι τέτοια ώστε να λαμβάνεται υπόψη το μέγεθος του αντίστοιχου πληθυσμού και το μέσο εισόδημα των κατοίκων της περιοχής.
- β) Του δειγματοληπτικού ή του πειραματικού σχεδιασμού. Η ανάγκη για συγκρότηση ομάδων ομοιογενών αντικειμένων μπορεί να οφείλεται στην εμπλοκή περισσότερων του ενός παρατηρητών ή πειραματιστών, στο διαφορετικό χρόνο και τόπο διεξαγωγής των επαναλήψεων της δειγματοληψίας ή του πειράματος, στη χρήση πολλαπλών οργάνων μέτρησης και στην ύπαρξη διαφορετικών μηχανισμών παραγωγής των δειγματοληπτικών (δειγματοληπτικό σχέδιο) ή πειραματικών δεδομένων (πειραματικό σχέδιο).

Σε κάθε περίπτωση, βασική επιδίωξη είναι τα αντικείμενα (οι δειγματοληπτικές ή οι πειραματικές μονάδες) μέσα σε κάθε ομάδα να είναι όσο το δυνατόν πιο ομοιογενή. Ενδεχομένως, όμως να παρουσιάζουν κάποιο είδος ανομοιογένειας μεταξύ των ομάδων. Με την ομαδοποίηση αναμένεται γενικά μείωση του δειγματοληπτικού ή πειραματικού σφάλματος.

Οι παρατηρήσεις - μετρήσεις επί των αντικειμένων μέσα σε κάθε ομάδα δεν μπορούν να θεωρηθούν ανεξάρτητες (Hox 1995, Goldstein 1991). Για παράδειγμα, στα γεωργικά πειράματα που διεξάγονται σε αγρούς, είναι δυνατό εξωγενείς παράγοντες, όπως η κλίση του εδάφους, τα θρεπτικά συστατικά και η υγρασία του, να επηρεάσουν τα αποτελέσματα των μετρήσεων και συνεπώς τα αντίστοιχα επαγωγικά συμπεράσματα (Steel & Torrie, 1986). Αν οι παράγοντες αυτοί δεν ληφθούν υπόψη κατά το σχεδιασμό του πειράματος, τότε η δράση τους θα είναι συσκοτισμένη με αυτή των πειραματικών επεμβάσεων. Οι κλασικοί στατιστικοί έλεγχοι υποθέσεων βασίζονται στην υπόθεση της ανεξαρτησίας των παρατηρήσεων. Αν αυτό δεν ισχύει, τότε τα αποτελέσματα των ελέγχων είναι αναξιόπιστα, ενώ το μέγεθος και η κατεύθυνση της μεροληψίας είναι απρόβλεπτη. Σε τέτοιες περιπτώσεις (μη ανεξαρτησίας), μάλιστα, η αύξηση του σφάλματος μπορεί να επέλθει ακόμη και όταν αυξάνει το μέγεθος του δείγματος. Με βάση τις παραπάνω επισημάνσεις, σε οποιαδήποτε σύγκριση ομάδων αντικειμένων θα πρέπει να ληφθεί υπόψη και να απομονωθεί η επίδραση της ομαδοποίησης στη μεταβλητότητα των υπό εξέταση παραμέτρων-μεταβλητών.

### 3. Εξειδίκευση του Προβλήματος

Έστω ότι έχουμε  $c$  ομάδες αντικειμένων, οι οποίες περιγράφονται από πίνακες της μορφής «αντικείμενα · μεταβλητές». Οι τιμές των μεταβλητών αντιστοιχούν στις διαθέσιμες μετρήσεις των παραμέτρων, οι οποίες αφορούν, για κάθε αντικείμενο, είτε σε διαφορετικές ιδιότητες είτε και στις ίδιες αλλά κάτω από διαφορετικές ή καταστάσεις στο χώρο και στο χρόνο. Από τη στιγμή που οι καταστάσεις αυτές αποτελούν διαφορετικές συνθήκες, οι οποίες έχουν επιλεγεί με βάση κάποιο προκαθορισμένο σχέδιο, όπου οι διαθέσιμες μετρήσεις θεωρούνται ότι εξαρτώνται από γνωστούς και ελεγχόμενους παράγοντες, τότε η όλη διαδικασία μπορεί να ενταχθεί στο πλαίσιο των πειραματικών σχεδιασμών. Χωρίς περιορισμό της γενικότητας, υποθέτουμε:

- α) ότι κάθε ομάδα αποτελείται από  $n$  αντικείμενα, τα οποία εξετάζονται ως προς  $p$  κοινές παραμέτρους-ποσοτικές μεταβλητές,
- β) ότι σε κάθε ομάδα οι τιμές των  $p$  παραμέτρων έχουν τυποποιηθεί ώστε σε κάθε περίπτωση να έχουν μέσο όρο ίσο με 0 (ανά στήλη) και
- γ) μία από τις  $c$  ομάδες, έστω η ομάδα  $t$ , θεωρείται ομάδα “στόχος-αναφοράς”. Έτσι, τα δεδομένα των ομάδων μπορούν να αποδοθούν με μια σειρά από  $c$  σε πλήθος  $n \cdot p$  πίνακες  $\mathbf{X}_i$  ( $i=1, \dots, c$ ) με στοιχεία από το σύνολο  $\hat{A}$ , των πραγματικών αριθμών.

Τέλος, το ειδικό πρόβλημα που πρέπει να επιλυθεί είναι το παρακάτω:

*Να κατασκευαστεί συνάρτηση-δείκτης που να μετρά το βαθμό προσέγγισης (ομοιότητας) ή απομάκρυνσης (ανομοιότητας) των  $c-1$  ομάδων από την ομάδα “στόχος”  $t$ , με βάση τις τιμές των  $p$  παραμέτρων-μεταβλητών, αφού προηγουμένως έχει απαλειφθεί από αυτές η όποια επίδραση έχει επιφέρει η ομαδοποίηση των αντικειμένων.*

### 4. Λύση του Προβλήματος

#### 4.1. Προκρούστια Μέθοδος

Λύση στο προηγούμενο πρόβλημα μπορεί να δώσει η «Προκρούστια Ορθογώνια Προβολή ή Περιστροφή - *Procrustes Orthogonal Projection or Rotation*». Σύμφωνα με τη μέθοδο αυτή, δοθέντων δύο  $m \cdot n$  πινάκων  $\mathbf{A}$  και  $\mathbf{B}$ , εκ των οποίων ο ένας, έστω ο  $\mathbf{B}$ , θεωρείται σταθερός, προσδιορίζεται  $n \cdot n$  πίνακας μετασχηματισμού  $\mathbf{T}$ , τέτοιος ώστε ο πίνακας  $\mathbf{AT}$  να προσεγγίζει βέλτιστα, με την έννοια των ελαχίστων τετραγώνων, τον  $\mathbf{B}$  (Gower & Dijk-

sterhuis 2004, Mardia, Kent & Bibby 2003, Golub & Van Loan 1989). Χωρίς περιορισμό της γενικότητας, θεωρούμε ότι οι στήλες των πινάκων  $\mathbf{A}$  και  $\mathbf{B}$  έχουν μέσο όρο ίσο με μηδέν. Σε αντιστοιχία με το πρόβλημα που εξετάζουμε, το ζητούμενο είναι οι χώροι στηλών των πινάκων  $\mathbf{X}_i$  να προβληθούν βέλτιστα, μετά από ορθογώνια περιστροφή, σε έναν κοινό χώρο “στόχο” – αναφοράς (*target or reference space*). Λόγω του γεγονότος ότι οι πίνακες  $\mathbf{X}_i$  έχουν τον ίδιο αριθμό γραμμών και στηλών, οι αντίστοιχοι χώροι έχουν την ίδια διάσταση. Επίσης, οι στήλες των πινάκων  $\mathbf{X}_i$  έχουν μέσο όρο ίσο με 0 και εκφράζουν τις αποκλίσεις των αντίστοιχων σημείων από το κέντρο βάρους τους. Επομένως, τα συστήματα συντεταγμένων των χώρων που ορίζονται από τις στήλες των  $\mathbf{X}_i$  έχουν κοινή αρχή. Συνεπώς, οποιοσδήποτε από τους  $c$  πίνακες  $\mathbf{X}_i$  μπορεί να χρησιμοποιηθεί ως χώρος αναφοράς, επί του οποίου θα προβληθούν οι υπόλοιποι  $c-1$  χώροι. Πιο συγκεκριμένα, έστω  $\mathbf{X}_t$  και  $\mathbf{X}_w$ , οι πίνακες με στοιχεία από την ομάδα στόχο-αναφοράς  $t$  και από την ομάδα  $w$  αντίστοιχα ( $t, w = 1, \dots, c \quad t \neq w$ ). Το πρόβλημα είναι να προσδιοριστεί πίνακας μετασχηματισμού  $\mathbf{R}$ , τέτοιος ώστε ο πίνακας  $\mathbf{X}_w \mathbf{R}$ , δηλαδή ο πίνακας με τα στοιχεία της ομάδας  $w$  μετά την περιστροφή μέσω του  $\mathbf{R}$ , να είναι όσο το δυνατόν πλησιέστερα στον πίνακα  $\mathbf{X}_t$ . Σύμφωνα με την Προκρούστια μέθοδο πρέπει να ελαχιστοποιηθεί η ποσότητα:

$$D = \text{trace}[(\mathbf{X}_t - \mathbf{X}_w \mathbf{R})^T (\mathbf{X}_t - \mathbf{X}_w \mathbf{R})],$$

όπου  $\mathbf{R}$  είναι ο ζητούμενος ορθογώνιος  $p \times p$  πίνακας περιστροφής με:

$$\mathbf{R}^T \mathbf{R} = \mathbf{R} \mathbf{R}^T = \mathbf{I}_p$$

Στην Ενότητα Α του Παραρτήματος αποδεικνύουμε ότι η  $D^{1/2}$  (συνεπώς και η  $D$ ) είναι μετρική.

Πρόκειται για ένα πρόβλημα ελαχίστων τετραγώνων και η λύση του επιτυγχάνεται με την παρακάτω διαδικασία (Andrade *et al.* 2004, Mardia, Kent & Bibby 2003, Krzanowski 1987, Rao 1980, Sibson 1978)<sup>1</sup>:

Αρχικά εφαρμόζεται ή μέθοδος SVD (Golub & Van Loan, 1989) στον πίνακα  $\mathbf{X}_w^T \mathbf{X}_t$ :

$$\mathbf{X}_w^T \mathbf{X}_t = \mathbf{U} \mathbf{D} \mathbf{V}^T.$$

1 Η διαδικασία στηρίζεται στο ότι το κλασικό πρόβλημα ελαχίστων τετραγώνων  $\min \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$  έχει δύο ισοδύναμες λύσεις (βλέπε Kalman, 1996):  $\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$  και  $\mathbf{x} = \mathbf{V} \mathbf{D}^{-1} \mathbf{U}^T \mathbf{b}$ .

Στη συνέχεια, ο πίνακας περιστροφής  $\mathbf{R}$  υπολογίζεται από τη σχέση:

$$\mathbf{R} = \mathbf{U}\mathbf{V}^T.$$

Μέσω της εφαρμογής της SVD στον πίνακα  $\mathbf{X}_i^T \mathbf{X}_i$  είναι δυνατό να επιτευχθεί περιστροφή, τάνυσμα ή/και ανάκλαση των στοιχείων του (Weller & Romney, 1990). Ο πίνακας μετασχηματισμού  $\mathbf{R} = \mathbf{U}\mathbf{V}^T$  περιστρέφει ή/και ανακλά τα στοιχεία του πίνακα  $\mathbf{X}_i$  στο χώρο στηλών του  $\mathbf{X}_i$  έτσι ώστε να ελαχιστοποιείται η ποσότητα  $D$ . Αν η προσαρμογή είναι τέλεια, τότε τα στοιχεία του πίνακα  $\mathbf{X}_i - \mathbf{X}_i \mathbf{R}$ , ο οποίος μετρά την απώλεια προσαρμογής μετά την περιστροφή, θα είναι ίσα με μηδέν με αποτέλεσμα και η μετρική  $D=0$ . Η ποσότητα  $D$  θεωρείται ως απόσταση ή ως δείκτης καλής προσαρμογής και η φυσική του ερμηνεία είναι ότι εκφράζει το βαθμό ομοιότητας των πινάκων  $\mathbf{X}_i$  και  $\mathbf{X}_i \mathbf{R}$ . Μια τυποποιημένη εκδοχή της απόστασης  $D$  αποτελεί ο δείκτης  $D^{**} = D / \text{trace}(\mathbf{X}_i \mathbf{X}_i^T)$  (Sibson, 1978).

Τα στοιχεία κάθε στήλης των πινάκων  $\mathbf{X}_i$  και  $\mathbf{X}_i \mathbf{R}$  είναι κεντροποιημένα ως προς τους μέσους όρους των στηλών και, επομένως, τα συστήματα συντεταγμένων των αντίστοιχων χώρων, οι οποίοι ορίζονται από τις στήλες τους, έχουν κοινή αρχή. Έτσι, δεν απαιτείται μεταφορά του συστήματος συντεταγμένων του χώρου που ορίζουν οι στήλες του πίνακα  $\mathbf{X}_i$ . Είναι δυνατό το νέφος σημείων του  $\mathbf{X}_i$  κατά την προβολή του στο χώρο αναφοράς  $\mathbf{X}_i$  να εμφανίζει διαφορετική διασπορά απ' ό,τι το νέφος των στοιχείων του  $\mathbf{X}_i$ , αφού κάθε ορθογώνιος μετασχηματισμός διατηρεί τις αποστάσεις (Groenen & Frances 2000, Peay 1988). Για να είναι συγκρίσιμα τα αποτελέσματα θα πρέπει να ληφθεί υπόψη και η διαφορετική μεταβλητότητα των αντίστοιχων νεφών (Mardia, Kent & Bibby 2003, Sibson 1978). Ένας απλός τρόπος για να αντιμετωπιστεί το ζήτημα είναι τα στοιχεία στηλών των πινάκων  $\mathbf{X}_i$  και  $\mathbf{X}_i \mathbf{R}$ , πριν την Προκρούστια Προβολή, να μετασχηματιστούν έτσι ώστε η διακύμανσή τους να είναι ίση με 1 (Krzanowski 1993, Peay 1988, Sibson 1978). Εναλλακτικά, στη γενική περίπτωση όπου τα στοιχεία των πινάκων  $\mathbf{X}_i$  και  $\mathbf{X}_i \mathbf{R}$  έχουν διαφορετικές κλίμακες μέτρησης, η συμβατότητα των δύο λύσεων μπορεί να επιτευχθεί με την επιβολή του περιορισμού τα στοιχεία του  $\mathbf{X}_i \mathbf{R}$ , μετά την περιστροφή, να ικανοποιούν τη σχέση:

$$z\mathbf{X}_i \mathbf{R}, \text{ με } z > 0.$$

Μπορεί να δειχθεί (Mardia, Kent & Bibby 2003, Sibson 1978) ότι η ελαχιστοποίηση της ποσότητας  $D$  επιτυγχάνεται θέτοντας:

$$z = \frac{[\text{trace}(\mathbf{D})]}{[\text{trace}(\mathbf{X}_w \mathbf{X}_w^T)]},$$

όπου  $\mathbf{D}$  είναι ο διαγώνιος πίνακας με στοιχεία τις χαρακτηριστικές τιμές που προκύπτουν από την εφαρμογή της SVD στον πίνακα  $\mathbf{X}_w^T \mathbf{X}_t$ . Με τον τρόπο αυτό, η κλίμακα μέτρησης των στοιχείων του  $\mathbf{X}_w$  μετασχηματίζεται σε αυτή των στοιχείων του  $\mathbf{X}_t$ . Θα πρέπει να παρατηρήσουμε ότι οι μετασχηματισμοί  $z\mathbf{X}_w \mathbf{R}$  και  $\mathbf{X}_w \mathbf{R}$  δεν είναι συμμετρικοί ως προς τις κλίμακες μέτρησης των στοιχείων των πινάκων  $\mathbf{X}_t$  και  $\mathbf{X}_w$  (Mardia, Kent & Bibby 2003, Sibson 1978, Gower 1975, Rummel 1970). Τα στοιχεία του  $\mathbf{X}_t$  δεν μπορούν να προκύψουν από τον αντίστροφο μετασχηματισμό των στοιχείων του  $\mathbf{X}_w$ . Η συμμετρία μπορεί να επιτευχθεί με τον καθορισμό μιας κοινής κλίμακας μέτρησης, τέτοιας ώστε τα αθροίσματα τετραγώνων των στοιχείων των  $\mathbf{X}_t$  και  $\mathbf{X}_w$  να είναι ίσα (Mardia, Kent & Bibby, 2003). Δηλαδή να ισχύει:

$$\text{trace}(\mathbf{X}_w \mathbf{X}_w^T) = \text{trace}(\mathbf{X}_t \mathbf{X}_t^T),$$

Στην πράξη, αφού επιλεγεί ο πίνακας αναφοράς  $\mathbf{X}_t$ , οι υπόλοιποι  $c-1$  πίνακες προβάλλονται, με την προτεινόμενη μεθοδολογία, στο χώρο που ορίζουν οι στήλες του  $\mathbf{X}_t$ .

#### 4.2. Διαγραμματικά και Αριθμητικά Αποτελέσματα

Στο πρόβλημα που εξετάζουμε το ενδιαφέρον εστιάζεται στη σύγκριση των  $n$  αντικειμένων μεταξύ των  $c$  ομάδων, αφού αφαιρεθεί η (ενδεχόμενη) επίδρασή τους στη διαμόρφωση των αποτελεσμάτων. Η οπτικοποίηση της όλης διαδικασίας επιτυγχάνεται επί κοινού χώρου (1, 2 ή 3 διαστάσεων). Η διαγραμματική ερμηνεία των αποτελεσμάτων στηρίζεται στη σύγκριση των σχετικών θέσεων των ομόλογων σημείων των αντικειμένων ή ομάδων τους επί του κοινού χώρου διαγραμματικής προβολής. Τα αριθμητικά αποτελέσματα της Προκρούστιας προβολής μπορούν να εμπλουτιστούν με τον υπολογισμό των παρακάτω δεικτών (Rummel, 1970):

- α) Των συντελεστών γραμμικής συσχέτισης του *Pearson* μεταξύ των στηλών των πινάκων  $\mathbf{X}_t$  και  $\mathbf{X}_w \mathbf{R}$ . Ιδιαίτερο ενδιαφέρον για την ποιότητα της προσρμογής παρουσιάζουν οι συντελεστές συσχέτισης των ομόλογων στηλών των ομάδων  $t$  (αναφοράς) και  $w$  μετά την περιστροφή. Οι τιμές τους εκφράζουν το βαθμό σύμπτωσης (*congruence*) των αντίστοιχων παραμέτρων-μεταβλητών.

- β) Των συντελεστών συσχέτισης (γραμμικής του *Pearson* και τάξεων του *Spearman*) μεταξύ των ομόλογων αντικειμένων στις ομάδες  $t$  και  $w$  μετά την περιστροφή. Οι αριθμητικές τιμές για τους υπολογισμούς βρίσκονται στις γραμμές των πινάκων  $\mathbf{X}_t$  και  $\mathbf{X}_w\mathbf{R}$ . Οι τιμές των συντελεστών αυτών εκφράζουν το βαθμό ομοιότητας των ομόλογων αντικειμένων στις δύο ομάδες. Χαμηλές τιμές των δεικτών υποδεικνύουν τα αντικείμενα που δέχθηκαν τη μεγαλύτερη επίδραση από την ομαδοποίηση. Βέβαια, το ίδιο μπορεί να διαπιστωθεί και από την εξέταση των στοιχείων του πίνακα  $\mathbf{X}_t - \mathbf{X}_w\mathbf{R}$ , ο οποίος περιέχει τις αποκλίσεις προσαρμογής ανά παράμετρο-μεταβλητή. Ο μέσος όρος ( $M\Sigma$ ) των συντελεστών συσχέτισης μεταξύ των ομόλογων αντικειμένων εκφράζει ένα συνολικό δείκτη καλής προσαρμογής της Προκρούστιας προβολής.
- γ) Των συντελεστών συσχέτισης (γραμμικής του *Pearson* και τάξεων του *Spearman*) μεταξύ των ομόλογων παραμέτρων στην ομάδα  $w$ , πριν και μετά την περιστροφή – προβολή. Για τον υπολογισμό τους χρησιμοποιούνται οι στήλες των πινάκων  $\mathbf{X}_w$  και  $\mathbf{X}_w\mathbf{R}$ . Οι πολύ χαμηλές τιμές και το αρνητικό πρόσημο των συντελεστών δηλώνουν μεταβολή της διάταξης των αντικειμένων και των παραμέτρων μετά την περιστροφή.

Η διαδικασία επαναλαμβάνεται για τις υπόλοιπες ομάδες. Στην Ενότητα Β του Παραρτήματος παρουσιάζονται τα αποτελέσματα εφαρμογής της προτεινόμενης μεθόδου σε πίνακες δεδομένων ενεργειακών εισροών-εκροών σε συμβατικά και βιολογικά και αγροοικοσυστήματα. Οι αναλύσεις πραγματοποιήθηκαν με το λογισμικό Procrustes v.1.0<sup>2</sup>.

#### 4.3. Παρατηρήσεις

- A) Με την Προκρούστια προβολή οι πίνακες  $\mathbf{X}_w$  καθίστανται όσο το δυνατόν πιο “όμοιοι” με τον πίνακα αναφοράς  $\mathbf{X}_t$ , αλλά και μεταξύ τους (Michailidis & De Leeuw 2000, Michailidis 1996). Εξομαλύνεται η μεταβλητότητα που εισάγει η ένταξη των αντικειμένων σε ομάδες και απαλείφεται η επίδρασή τους από τις τιμές των παρα-

---

2 Το συγκεκριμένο λογισμικό κατασκευάστηκε από τους Α. Μάρκο και Γ. Μενεξέ στο πλαίσιο του προγράμματος με τίτλο “Δημιουργία Δεικτών για τη Βελτίωση της Μέτρησης Κοινωνικο-Οικονομικών Παραμέτρων (ΔΕ-ΚΟ-ΠΑ)” με επιστημονικό υπεύθυνο τον Ν. Φαρμάκη.

μέτρων (Rummel, 1970).

- B)** Η μέθοδος που παρουσιάσαμε δεν είναι εφικτό να υλοποιηθεί άμεσα αν οι πίνακες  $X_i$  δεν έχουν το ίδιο πλήθος γραμμών<sup>3</sup>, δηλαδή όταν οι ομάδες δεν περιέχουν τον ίδιο αριθμό αντικειμένων. Στην περίπτωση αυτή, αν για παράδειγμα ο πίνακας  $X_{11}$  έχει λιγότερες γραμμές από τον  $X_1$ , τότε οι υπολειπόμενες γραμμές του  $X_{11}$  μπορούν να συμπληρωθούν με μηδενικά, ώστε τελικά οι δύο πίνακες να έχουν τον ίδιο αριθμό γραμμών (Mardia, Kent & Bibby, 2003).

### 5. Συζήτηση-Συμπεράσματα

Με την Προκρούστια προβολή – περιστροφή τα αντικείμενα εντός των ομάδων προβάλλονται σε κοινό χώρο και οι σχετικές τους θέσεις είναι συγκρίσιμες. Φυσικά, το ίδιο ισχύει και για τις τιμές των παραμέτρων-μεταβλητών. Την ίδια μέθοδο περιστροφής προτείνουν οι Michailidis (1996) και Michailidis και De Leeuw (2000) για τη σύγκριση των αποτελεσμάτων που προκύπτουν από την εφαρμογή της Ανάλυσης Ομοιογένειας σε  $c$  συστάδες δεδομένων, τα οποία έχουν συγκεντρωθεί με το δειγματοληπτικό σχήμα των Πολλαπλών Επιπέδων (*Multilevel*). Σε αντιστοιχία με την περίπτωση του προβλήματος που εξετάζουμε οι  $c$  συστάδες μπορούν να θεωρηθούν ως ομάδες. Επίσης, η μέθοδος χρησιμοποιείται για τη σύγκριση των υποχώρων που προκύπτουν από την εφαρμογή της Παραγοντικής Ανάλυσης (Raykov & Little 1999, Gruvaeus 1970, Harman 1967) και της Πολυδιάστατης Κλιμάκωσης (Peay 1988, Sibson 1978) σε δύο ή περισσότερα ανεξάρτητα σύνολα δεδομένων. Βρίσκει εφαρμογή στους ελέγχους εξωτερικής και εσωτερικής σταθερότητας των αποτελεσμάτων της Παραγοντικής Ανάλυσης των Αντιστοιχιών όταν αυτοί πραγματοποιούνται με μεθόδους επαναδειγματοληψίας, όπως είναι η *Bootstrap* (Greenacre 2006, Markus 1994α και 1994β), καθώς και σε μεθοδολογίες για την επιλογή μεταβλητών στην Ανάλυση σε Κύριες Συνιστώσες (Andrade *et al.* 2004, Krzanowski 1987). Γενικεύσεις και επιπλέον εφαρμογές της Προκρούστιας Προβολής συναντάμε στους Schönemann (1966), Rummel (1970), Gower (1975), Sibson (1978), Lebart, Morineau και Piron (2000), Gower και Dijksterhuis (2004) και Gardner, Gower και Le Roux (2006).

---

3 Πιο συγκεκριμένα, δεν μπορεί να εφαρμοστεί άμεσα η Προκρούστια Προβολή, αφού οι πίνακες δεν έχουν το ίδιο μέγεθος.

Συμπερασματικά, με την προτεινόμενη μεθοδολογία επιτυγχάνονται τα εξής:

- 1) απομονώνεται η επίδραση της ομαδοποίησης των αντικειμένων στις τιμές των παραμέτρων
- 2) οι ομάδες ομογενοποιούνται, είναι συγκρίσιμες και επιπλέον αναδεικνύονται διαγραμματικά οι μεταξύ τους διαφοροποιήσεις, ως προς τις σχέσεις των προφίλ των αντικειμένων, που ενθυλακώνουν οι πίνακες που αναλύονται. Ως μέτρο ομοιότητας μεταξύ της ομάδας αναφοράς και των προβαλλόμενων επί αυτής ομάδων μπορεί να χρησιμοποιηθεί η Προκρούστια απόσταση  $D$  (ή  $D^{**}$ ).

### Παράρτημα Α

#### Απόδειξη ότι η $D^{1/2}$ είναι μετρική

##### 1° Αξίωμα

Προφανώς,  $D^{1/2} \geq 0$ , ως θετική τετραγωνική ρίζα αθροίσματος τετραγώνων διαφορών.

##### 2° Αξίωμα

Έστω ότι έχουμε  $2 \cdot n \cdot p$  πίνακες  $\mathbf{X}_1$  και  $\mathbf{X}_2$ . Χωρίς περιορισμό της γενικότητας μπορούμε να θεωρήσουμε ότι τα  $np$  στοιχεία κάθε πίνακα είναι διατεταγμένα ως προς τις στήλες του πίνακα και ότι αποτελούν τις συντεταγμένες ενός σημείου σε έναν Ευκλείδειο  $np$ -χώρο. Έτσι, οι 2 πίνακες  $\mathbf{X}_1$  και  $\mathbf{X}_2$  αναπαρίστανται από 2 σημεία  $Q_1$  και  $Q_2$  αντίστοιχα. Η ευκλείδεια απόσταση  $\Delta$  (όχι απαραίτητα η ελάχιστη) μεταξύ των σημείων  $Q_1$  και  $Q_2$  είναι η θετική τετραγωνική ρίζα του αθροίσματος τετραγώνων των διαφορών των συντεταγμένων τους και αποτελεί ένα μέτρο της καλής προσαρμογής (ή της απόστασης) μεταξύ των πινάκων  $\mathbf{X}_1$  και  $\mathbf{X}_2$ . Αν τώρα ο πίνακας  $\mathbf{X}_1$  έχει περιστραφεί (μέσω της Προκρούστιας μεθόδου) ώστε να έχει τη μεγαλύτερη δυνατή ομοιότητα με τον πίνακα  $\mathbf{X}_2$  και ο πίνακας  $\mathbf{X}_2$  έχει περιστραφεί ώστε να έχει τη μεγαλύτερη δυνατή ομοιότητα με τον πίνακα  $\mathbf{X}_1$  τότε οι ποσότητες  $D_{12}^{1/2}$  και  $D_{21}^{1/2}$  αντιστοιχούν στις αποστάσεις  $\Delta(Q_1, Q_2)$  και  $\Delta(Q_2, Q_1)$ . Όμως, ισχύει:

- i)  $D_{12}^{1/2} \geq D_{21}^{1/2}$  (γιατί ο πίνακας  $\mathbf{X}_1$  δεν μπορεί να περιστραφεί με καλύτερο τρόπο από αυτόν της προκρούστιας μεθόδου ώστε η ποσότητα  $D_{12}$  να είναι ελάχιστη)

ii) ( $D_{21}^{1/2} \leq D_{12}^{1/2}$ ) (γιατί ο πίνακας  $\mathbf{X}_2$  δεν μπορεί να περιστραφεί με καλύτερο τρόπο από αυτόν της προκρούστιας μεθόδου ώστε η ποσότητα  $D_{21}$  να είναι ελάχιστη)

Από τις i) και ii) και με δεδομένο ότι  $D_{12}^{1/2}, D_{21}^{1/2} \geq 0$ , προκύπτει:

$$D_{12}^{1/2} = D_{21}^{1/2}.$$

### 3° Αξίωμα

Έστω ότι έχουμε 3  $n \cdot p$  πίνακες  $\mathbf{X}_1, \mathbf{X}_2$  και  $\mathbf{X}_3$ . Χωρίς περιορισμό της γενικότητας μπορούμε να θεωρήσουμε ότι τα  $np$  στοιχεία κάθε πίνακα είναι διατεταγμένα ως προς τις στήλες του πίνακα και ότι αποτελούν τις συντεταγμένες ενός σημείου σε έναν Ευκλείδειο  $np$ -χώρο. Έτσι, οι 3 πίνακες  $\mathbf{X}_1, \mathbf{X}_2$  και  $\mathbf{X}_3$  αναπαρίστανται από 3 σημεία  $Q_1, Q_2$  και  $Q_3$  αντίστοιχα. Η ευκλείδεια απόσταση  $\Delta$  (όχι απαραίτητα η ελάχιστη) μεταξύ δύο οποιονδήποτε σημείων  $Q_i$  και  $Q_j$  ( $i, j=1, 2, 3$ ) είναι η θετική τετραγωνική ρίζα του αθροίσματος τετραγώνων των διαφορών των συντεταγμένων τους και αποτελεί ένα μέτρο της καλής προσαρμογής (ή της απόστασης) μεταξύ των πινάκων  $\mathbf{X}_i$  και  $\mathbf{X}_j$ . Οι ποσότητες  $D_{13}^{1/2}$  και  $D_{23}^{1/2}$  αντιστοιχούν στις αποστάσεις  $\Delta(Q_1, Q_3)$  και  $\Delta(Q_2, Q_3)$  με πίνακες περιστροφής αντίστοιχα  $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{I}$ . Αν τώρα οι πίνακες  $\mathbf{X}_1$  και  $\mathbf{X}_2$  έχουν περιστραφεί (μέσω της Προκρούστιας μεθόδου) ώστε να έχουν τη μεγαλύτερη δυνατή ομοιότητα με τον πίνακα  $\mathbf{X}_3$ , τότε οι ποσότητες  $D_{13}^{1/2}$  και  $D_{23}^{1/2}$  αντιστοιχούν στις αποστάσεις  $\Delta(Q_1, Q_3)$  και  $\Delta(Q_2, Q_3)$  με πίνακες περιστροφής  $\mathbf{R}_1 \neq \mathbf{R}_2$ . Επειδή αυτά συμβαίνουν σε Ευκλείδειο χώρο ισχύει:

$$\Delta(Q_1, Q_2) \leq D_{13}^{1/2} + D_{23}^{1/2}.$$

Όμως,  $\Delta(Q_1, Q_2) \geq D_{12}^{1/2}$ , γιατί οι πίνακες  $\mathbf{X}_1$  και  $\mathbf{X}_2$  δεν μπορούν να περιστραφούν με καλύτερο τρόπο από αυτόν της προκρούστιας μεθόδου ώστε η ποσότητα  $D_{12}$  να είναι ελάχιστη. Επομένως ισχύει η ανισότητα:

$$D_{12}^{1/2} \leq D_{13}^{1/2} + D_{23}^{1/2}.$$

## Παράρτημα Β

### Παράδειγμα Εφαρμογής

Η προτεινόμενη μεθοδολογία εφαρμόστηκε σε 5 πίνακες δεδομένων 4.3. Κάθε πίνακας αντιστοιχεί σε μία τοποθεσία όπου μετρήθηκε το σύνολο των ενεργειακών εκροών (Y1) και το σύνολο των ανανεώσιμων (Y2) και μη ενεργειακών εισροών (Y3) σε δύο βιολογικά και σε δύο συμβατικά αγρο-

οικοσυστήματα (4 αντικείμενα). Τα δεδομένα δίνονται στον Πίνακα 1. Με βάση αντικειμενικά κριτήρια διαχείρισης ενεργειακών εισροών και εκροών για τα συγκεκριμένα αγροοικοσυστήματα προσδιορίστηκε ο ιδεατός πίνακας “στόχος” (βλέπε Πίνακα 1). Στο Διάγραμμα 1 οπτικοποιείται η σύγκλιση των 5 τοποθεσιών ως προς τον πίνακα “στόχο”. Από το Διάγραμμα 1 διαπιστώνεται ότι πιο κοντά (σύγκλιση) στον πίνακα στόχο βρίσκεται η τοποθεσία 1 (X1) και πιο μακριά (απόκλιση) η τοποθεσία 3 (X3). Οι τοποθεσίες 2 (X2), 4 (X4) και 5 (X5) βρίσκονται κοντά μεταξύ τους και περίπου στη μέση της απόστασης μεταξύ των δύο ακραίων τοποθεσιών 1 και 3. Στον Πίνακα 2 παρουσιάζονται οι αντίστοιχες προκρούστιες αποστάσεις. Οι συντελεστές συσχέτισης (Pearson) μεταξύ των ομόλογων στηλών των πινάκων  $X_i$  και  $X_wR$  κυμάνθηκαν από 0,810 έως 0,991 και οι συντελεστές συσχέτισης μεταξύ των ομόλογων γραμμών των παραπάνω πινάκων από 0,898 έως 1,000. Οι τιμές αυτές δηλώνουν υψηλό βαθμό σύμπτωσης των μεταβλητών και των αντικειμένων των πέντε πινάκων μετά την περιστροφή. Από τη μελέτη των συντελεστών συσχέτισης των ομόλογων στηλών των πινάκων  $X_w$  και  $X_wR$  διαπιστώθηκε ότι αρνητική συσχέτιση εμφανίζεται για τη μεταβλητή Y1 (ενεργειακές εκροές) της τρίτης τοποθεσίας ( $r=-0,101$ ). Το αποτέλεσμα αυτό δηλώνει ότι οι τιμές των 4 αγροοικοσυστημάτων στη μεταβλητή Y1 επηρεάστηκαν από τα ιδιαίτερα χαρακτηριστικά της τρίτης τοποθεσίας, γεγονός που χρήζει περαιτέρω διερεύνησης.

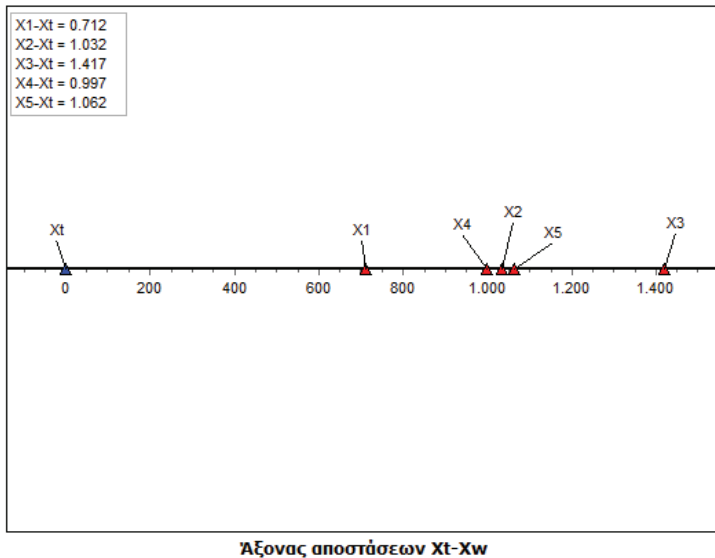
Πίνακας 1: Ενεργειακές Εισροές και Εκροές των Αγροοικοσυστημάτων

Τοποθεσίες	Αγροοικοσυστήματα	Y1 Εκροές (GJ/ha)	Y2 Ανανεώσιμες Εισροές (MJ/ha)	Y3 Μη Ανανεώσιμες Εισροές (MJ/ha)
Τοποθεσία 1 (πίνακας X1)	Βιολογικό 1	28,129	35287,0	32377,6
	Βιολογικό 2	37,329	25861,3	33983,6
	Συμβατικό 1	40,880	9973,9	45480,3
	Συμβατικό 2	47,320	9423,0	44779,3
Τοποθεσία 2 (πίνακας X2)	Βιολογικό 1	24,711	14932,6	29668,4
	Βιολογικό 2	32,071	41553,6	29624,1
	Συμβατικό 1	50,475	14437,1	46374,3
	Συμβατικό 2	50,475	7565,7	47765,2
Τοποθεσία 3 (πίνακας X3)	Βιολογικό 1	34,175	24182,7	31238,3
	Βιολογικό 2	34,175	14111,3	31319,1
	Συμβατικό 1	52,315	5567,0	44107,7
	Συμβατικό 2	33,915	8347,0	48661,7
Τοποθεσία 4	Βιολογικό 1	24,975	40732,3	34219,1

(πίνακας Χ4)	Βιολογικό 2	24,975	34465,7	29919,6
	Συμβατικό 1	50,475	11900,5	44151,6
	Συμβατικό 2	50,475	11005,7	45837,3
Τοποθεσία 5	Βιολογικό 1	23,134	3131,3	31459,4
(πίνακας Χ5)	Βιολογικό 2	23,134	51922,8	31515,4
	Συμβατικό 1	38,120	11075,7	42262,8
	Συμβατικό 2	38,120	9109,1	42207,2
Στόχος	Βιολογικό 1	30	35000	35000
(πίνακας Χt)	Βιολογικό 2	40	30000	30000
	Συμβατικό 1	41	10000	45000
	Συμβατικό 2	47	9500	44000

Πίνακας 2: Προκρούστιες αποστάσεις των πέντε πινάκων από τον πίνακα “στόχο”

	X1-Xt	X2-Xt	X3-Xt	X4-Xt	X5-Xt
$D$	0,507	1,066	2,007	0,993	1,127
$D^{1/2}$	0,712	1,032	1,417	0,997	1,062



Διάγραμμα 1: Οπτικοποίηση της σύγκλισης των 5 πινάκων (τοποθεσιών) ως προς τον ιδεατό πίνακα “στόχο”.

## Ευχαριστίες

Θέλουμε να εκφράσουμε τις θερμές μας ευχαριστίες στον Επ. Καθηγητή της Γεωπονική Σχολής του ΑΠΘ κ. Ανδρέα Μαμάλο για την παραχώρηση των δεδομένων που χρησιμοποιήθηκαν στο παράδειγμα εφαρμογής της προτεινόμενης μεθοδολογίας.

# Index for measuring the convergence between groups of data and a target (reference) data table: Application to energy inputs-outputs tabled data of selected agro-ecosystems

George Menexes<sup>1</sup>, Angelos Markos<sup>2</sup>, Nicolas Farmakis<sup>3</sup>

<sup>1</sup>Lab of Agronomy, School of Agriculture, Aristotle University of Thessaloniki, Greece, [gmenexes@agro.auth.gr](mailto:gmenexes@agro.auth.gr)

<sup>2</sup>Department of Primary Education, Lab of Mathematics & Informatics, [amarkos@eled.duth.gr](mailto:amarkos@eled.duth.gr)

<sup>3</sup>Dept of Mathematics, Aristotle University of Thessaloniki, Greece [farmakis@math.auth.gr](mailto:farmakis@math.auth.gr)

## Abstract

In some practical applications there is a need to compare two or more group of objects characterized by some common attributes (parameters) with a similarly structured target group. The values of the parameters of the target group correspond to comparison criteria or comprise theoretical-ideal values useful for the evaluation of the degree of convergence of the other groups towards the target group. In each case, these target values are determined by the corresponding scientific research field. Suppose that we have  $c$  groups of  $n$  objects that are described by data tables  $\mathbf{X}_{i(n,p)}$  of the form “objects’ variables”. The  $p$  variables are quantitative, measured on interval or ratio scales, and their values correspond to measurements of some objects’ attributes, in same or different situations in space or time. In this study, we attack the following general problem: find a function–index that measures the degree of the convergence or similarity of the  $c-1$  objects groups from a target group  $t$ , based on the values of the  $p$  variables after eliminating the effect of grouping of

the objects. A solution could be established by means of the Procrustes orthogonal projection-rotation. The proposed method is applied to energy inputs-outputs tabulated data of selected agro-ecosystems.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- Andrade, J. M., Gómez – Carracedo, M., Krzanowski, W. & Kubista, M. (2004). Procrustes Rotation in Analytical Chemistry, a Tutorial. *Chemometrics and Intelligent Laboratory Systems*, **72**, 123-132.
- Dijksterhuis, G., Martens, H. & Martens, M. (2005). Combined Procrustes Analysis and PLSR for Internal and External Mapping of Data from Multiple Sources. *Computational Statistics & Data Analysis*, **48**, 47-62.
- Gardner, S. Gower, J. & Le Roux, N. J. (2006). A Synthesis of Canonical Variate Analysis, Generalized Canonical Correlation and Procrustes Analysis. *Computational Statistics & Data Analysis*, **50**, 107-134.
- Goldstein, H. (1991). Multilevel Modelling of Survey Data. *The Statistician*, **40**(2), Special Issue: Survey Design, Methodology and Analysis, 235-244.
- Golub, G. & Van Loan, C. (1989). *Matrix Computations*, 2<sup>nd</sup> edition. Baltimore and London: The Johns Hopkins University Press.
- Goodal, C. (1991). Procrustes Methods in the Statistical Analysis of Shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, **53**(2), 285-339.
- Gower, J. (1975). Generalized Procrustes Analysis. *Psychometrika*, **40**(1), 33-51.
- Gower, J. & Dijksterhuis, G. (2004). *Procrustes Problems*. Oxford Statistical Science Series, 30. Oxford University Press.
- Greenacre, M. (2006). Tying Up the Loose Ends in Simple Correspondence Analysis. *Economics Working Papers 940*, Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain.
- Groenen, P. & Franses, P. (2000). Visualizing Time-Varying Correlations Across Stock Market. *Journal of Empirical Finance*, **7**, 155-172.
- Gruvaeus, G. (1970). A General Approach to Procrustes Pattern Rotation. *Psychometrika*, **35**(4), 493-505.
- Harman, H. H. (1967). *Modern Factor Analysis*. Chicago: University of Chicago Press.
- Hox, J. J. (1995). *Applied Multilevel Analysis*. Amsterdam: TT-Publikaties.
- Kirk, R. (1995). *Experimental Design: Procedures for the Behavioral Sciences*. Pacific Grove: Brooks/Cole Publishing Company.

- Krzanowski, W. J. (1987). Selection of Variables to Preserve Multivariate Data Structure, Using Principal Components. *Applied Statistics*, 36(1), 22-33.
- Krzanowski, W. J. (1993). Attribute Selection in Correspondence Analysis of Incidence Matrices. *Applied Statistics*, 42(3), 529-541.
- Kutner, M., Nachtsheim, C., Neter, J. & Li, W. (2005). *Applied Linear Statistical Models*. Singapore: McGraw-Hill, Inc.
- Lebart, L., Morineau, A. & Piron, M. (2000). *Statistique Exploratoire Multidimensionnelle*. Paris: Dunod.
- Mardia, K., Kent, J. & Bibby, J. (2003). *Multivariate Analysis*. London: Academic Press.
- Markus, M. (1994α). *Bootstrap Confidence Regions in Nonlinear Multivariate Analysis*. Leiden University Leiden: DSWO Press.
- Markus, M. (1994β). Bootstrap Confidence Regions for Homogeneity Analysis; the Influence of Rotation on Coverage Percentages. In R. Dutter and W. Grossmann (Eds), *Proceedings: Computational Statistics COMPSTAT*, (pp. 337-342).
- Michailidis, G. & De Leeuw, J. (2000). Multilevel Homogeneity Analysis With Differential Weighting. *Computational Statistics & Data Analysis*, 32, 411-442.
- Michailidis, G. (1996). *Multilevel Homogeneity Analysis*. Διδακτορική διατριβή που υποβλήθηκε στο Τμήμα Μαθηματικών του Πανεπιστημίου της Καλιφόρνια στο Λος Άντζελες.
- Peay, E. (1988). Multidimensional Rotation and Scaling of Configurations to Optimal Agreement. *Psychometrika*, 53(2), 199-208.
- Rao, C. R. (1980). Matrix Approximations and Reduction of Dimensionality in Multivariate Analysis. In P. R. Krishnaiah (Ed.), *Multivariate Analysis – V, Proceedings of the Fifth International Symposium on Multivariate Analysis*, (pp. 3-22). Amsterdam: North – Holland Publishing Company.
- Raykov, T. & Little, T. (1999). A Note on Procrustean Rotation in Exploratory Factor Analysis: A Computer Intensive Approach to Goodness-of-Fit Evaluation. *Educational and Psychological Measurement*, 59(1), 47-57.
- Rummel, R. J. (1970). *Applied Factor Analysis*. Evanston: Northwestern University Press.
- Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Willey & Sons, Inc.
- Sibson, R. (1978). Studies in the Robustness of Multidimensional Scaling: Procrustes Statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(2), 234-238.
- Schönemann, P. (1966). A Generalized Solution of the Orthogonal Procrustes Problem. *Psychometrika*, 31(1), 1-10.

Steel, R. & Torrie, J. (1986). *Principles and Procedures of Statistics: A Biometrical Approach*. Singapore: McGraw-Hill Book Company.

Weller, S. & Romney A. K. (1990). *Metric Scaling: Correspondence Analysis*. Newbury Park: Sage Publications.