

Ανάλυση Δεδομένων (ΑΔ) και Στατιστική: Διαλεκτική και Συμπληρωματικότητα

Γεώργιος Μενεξές

Γεωπονική Σχολή ΑΠΘ, Εργαστήριο Γεωργίας

Περίληψη

Η Ανάλυση Δεδομένων (ΑΔ) αποτελεί κλάδο της ενότητας των μεθόδων της Πολυδιάστατης Στατιστικής Ανάλυσης και εκφράζει μια εναλλακτική μεθοδολογική και φιλοσοφική προσέγγιση στη στατιστική συμπερασματολογία, η οποία έρχεται σε αντίθεση ή ακόμη και σε ρήξη με την κλασική αγγλοσαξονική παράδοση του στατιστικού ελέγχου υποθέσεων. Στην παρούσα εργασία, πρώτον επιχειρείται κριτική οριοθέτηση του εννοιολογικού πεδίου της ΑΔ, όπως αυτό καθορίζεται στο επιστημολογικό πλαίσιο των δύο σημαντικότερων Σχολών ΑΔ (Γαλλικής και Ολλανδικής) και δεύτερον αναδεικνύεται η σχέση και η συμπληρωματικότητά της με τη Στατιστική γενικότερα.

Data Analysis and Statistics: Dialectic and Complementarity

George Menexes

School of Agriculture, Lab of Agronomy, Aristotle University of Thessaloniki

Abstract

Data Analysis is a branch of the Multidimensional Statistical Analysis methods and supports an alternative methodological and philosophical approach to inductive reasoning. This approach contradicts to the traditional statistical hypothesis testing procedures. The objectives of this paper are, first, to offer a critical review of the Data Analysis' conceptual field within the epistemological frame of two outstanding schools of Data Analysis (French and Dutch) and, secondly, to reveal its complementary relation to Statistics.

1. Η Ανάλυση Δεδομένων

Η Ανάλυση Δεδομένων (*Data Analysis* στα αγγλικά και *L'Analyse des Données* στα γαλλικά) αποτελεί κλάδο της ενότητας των μεθόδων της Πολυδιάστατης Στατιστικής Ανάλυσης (Καραπιστόλης 1999, Μπεχράκης 1999, Παπαδημητρίου 2007) και περιλαμβάνει, σύμφωνα με τους Deville και Malinvaud (1983), τρεις βασικές οικογένειες μεθόδων: α) την Παραγοντική Ανάλυση των Αντιστοιχιών-ΠΑΑ (διμεταβλητή και πολυμεταβλητή), β) την Ανάλυση σε Κύριες Συνιστώσες και γ) την Ταξινόμηση σε Αύξουσα Ιεραρχία. Ιδιαίτερο χαρακτηριστικό των μεθόδων αυτών είναι η συμμετρική αντιμετώπιση των μεταβλητών, όπου δεν υπάρχει διάκριση μεταξύ εξαρτημένων και ανεξάρτητων. Άλλες γνωστές και διαδεδομένες σε εφαρμογή μέθοδοι της ΑΔ είναι η Παραγοντική Ανάλυση, η Ανάλυση Κανονικοποιημένης Συσχέτισης, η Πολυδιάστατη Κλιμακοποίηση και η Διακρίνουσα Ανάλυση (Dillon & Goldstein 1984, Johnson & Wichern 1992, Hair *et al.* 1995, Sharma 1996, Tacc 1997, Johnson 1998, Stevens 2002). Σκοπός των μεθόδων είναι να αναδείξουν και να περιγράψουν λανθάνουσες δομές που ενδεχομένως εμπεριέχονται σε πολυδιάστατους πίνακες δεδομένων. Αυτό επιτυγχάνεται μέσα από διαδικασίες αλλαγής και ελάττωσης των διαστάσεων του αρχικού χώρου, στον οποίο το υπό εξέταση φαινόμενο μπορεί να περιγραφεί. Οι νέες διαστάσεις, οι οποίες δομούνται συνήθως από πολύπλοκες σχέσεις μεταξύ των διαθέσιμων μετρήσεων, ερμηνεύονται τελικά ως νέες σύνθετες μεταβλητές ή παράγοντες

(Dillon & Goldstein 1984, Παπαδημητρίου 1994). Οι μέθοδοι, σε ένα πρώτο επίπεδο, δεν απαιτούν την *a priori* παραδοχή ύπαρξης κάποιας θεωρητικής κατανομής ή κάποια υπόθεση σχετικά με τις παραμέτρους του υπό εξέταση πληθυσμού ή πληθυσμών, δηλαδή την ύπαρξη κάποιου στοχαστικού υποδείγματος (Benzécri, 1991). Σύμφωνα με τον Καραπιστόλη (1999, σ.21), “*Η ανάγκη λοιπόν να μη θεωρείται εκ των προτέρων ότι ένα φαινόμενο ακολουθεί κάποιο συγκεκριμένο νόμο, οδήγησε στην εφαρμογή νέων στατιστικών μη παραμετρικών μεθόδων, κάτω από την ονομασία Ανάλυση Δεδομένων ή όπως αλλιώς μπορεί να την αποκαλέσουμε Στατιστική δίχως μοντέλα.*” Κάτω από αυτή τη θεώρηση, η ΑΔ φαίνεται να εκφράζει μια νέα προσέγγιση στη στατιστική συμπερασματολογία, η οποία έρχεται σε αντίθεση με την αγγλοσαξονική παράδοση των στατιστικών ελέγχων (Αθανασιάδης, 1995). Μάλιστα, ο Gras (1995), φθάνοντας ίσως στην υπερβολή, μιλάει τελικά για επιστημολογική ρήξη της ΑΔ με την Κλασική Στατιστική.

Για τους Αγγλοσάξονες ο όρος “Ανάλυση Δεδομένων” δηλώνει μια προσέγγιση των στατιστικών αναλύσεων με ιδιαίτερη προσήλωση στα δεδομένα καθαυτά, αφήνοντας, προσωρινά τουλάχιστον, στο περιθώριο το θεωρητικό πλαίσιο που οδήγησε στη συγκρότησή τους (Αθανασιάδης, 1995). Η οπτική αυτή οδήγησε στην ανάπτυξη της Πολυδιάστατης Κλιμακοποίησης (Kruskal & Shepard 1974, Kruskal & Wish 1978, Hair *et al.* 1995) στις Η.Π.Α. και της μεθόδου της ΠΑΑ στη Γαλλία, κάτω από την καταλυτική επίδραση του Jean-Paul Benzécri (Deville & Malinvaud 1983, Greenacre 1993, Van Meter *et al.* 1994, Καραπιστόλης 1999, Μπεχράκης 1999, Παπαδημητρίου 2002 και 1994, Le Roux & Rouanet 2004). Για τον Benzécri η ΑΔ συνιστά Φιλοσοφία, η οποία απελευθερώνει τον ερευνητή από δεσμεύσεις, που ενδεχομένως επιβάλλουν εξωγενείς, σε σχέση με την έρευνα, παράγοντες, αφήνοντάς του τη φροντίδα και την ευθύνη να εξάγει ο ίδιος τις ερμηνείες των φαινομένων και τις συνέπειές τους (Benzécri & *Collaborateurs*, 1973).

Ο σημαντικότερος, ίσως, παράγοντας που οδήγησε στην υιοθέτηση των μεθόδων της ΑΔ από ένα ευρύ φάσμα επιστημονικών πεδίων ήταν η ανάπτυξη και η γενικευμένη χρήση των Ηλεκτρονικών Υπολογιστών (Η/Υ) (Μπεχράκης 1999, Παπαδημητρίου 2002). Οι μέθοδοι απαιτούν πολύπλοκους αριθμητικούς υπολογισμούς που μόνο με τη βοήθεια Η/Υ είναι δυνατό, μέσα στα όρια της ανθρώπινης υπομονής, να επιτευχθούν, ιδιαίτερα όταν πρόκειται να αναλυθούν μεγάλα σύνολα δεδομένων. Στις μέρες μας, δημοφιλή εμπορικά στατιστικά πακέτα όπως το SAS (SAS Institute, 1999), το BMDP (Moran & Gornbein 1988, BMDP Inc. 1992) και το SPSS (Norusis 1992, Meulman & Heiser 2004), περιλαμβάνουν μεθόδους όπως η ΠΑΑ, η Ανάλυση σε Κύριες Συνιστώσες και η Ιεραρχική Ταξινόμηση. Ειδικότερα, η ΠΑΑ αποτελεί τον κορμό πάνω στον οποίο στηρίζονται οι μεθοδολογικές προσεγγίσεις των δύο σημαντικότερων «Σχολών» ΑΔ, της Γαλλικής και της Ολλανδικής.

2. Σχολές Ανάλυσης Δεδομένων

Πριν αναφερθούμε στις Σχολές ΑΔ, κρίνουμε σκόπιμο να παραθέσουμε μια σύντομη ιστορική ανασκόπηση σχετικά με την ανάπτυξη της ΠΑΑ. Κίνητρο για την αναδρομή αυτή δεν είναι μόνο η ικανοποίηση μιας νόμιμης ιστοριογραφικής περιέργειας, αλλά και η πεποίθηση ότι η παρακολούθηση της εσωτερικής εξέλιξης της μεθόδου συμβάλλει, ως ένα βαθμό, στην αιτιολόγηση της χρησιμότητας και της αναγκαιότητάς της τόσο σε πρακτικό όσο και σε θεωρητικό επίπεδο. Για περισσότερα στοιχεία, σχετικά με την ιστορική εξέλιξη της μεθόδου, παραπέμπουμε στους Van Rijckevorsel (1987), Greenacre (1984), Tenenhaus και Young (1985), Van Rijckevorsel και De Leeuw (1988), Van Meter *et al.* (1994), Nishisato (1996 και 1980), Clausen (1998), Beh (2004) και Le Roux και Rouanet (2004).

Κατά τη διάρκεια του 20^{ου} αιώνα η ΠΑΑ εμφανίστηκε και αναπτύχθηκε ανεξάρτητα και, σε ορισμένες περιπτώσεις, σχεδόν ταυτόχρονα σε αρκετές χώρες, όπως οι Η.Π.Α., η Μεγάλη Βρετανία, ο Καναδάς, η Γαλλία, η Ολλανδία και η Ιαπωνία (Clausen, 1998). Αυτή η

παράλληλη ανάπτυξη είχε ως αποτέλεσμα να δημιουργηθούν διάφορες προσεγγίσεις και κατευθύνσεις (Σχολές), τόσο ως προς το θεωρητικό όσο και ως προς το αλγοριθμικό – υπολογιστικό υπόβαθρο της μεθόδου. Έτσι, η μέθοδος έγινε γνωστή με διαφορετικά ονόματα όπως Δυϊκή Κλιμακοποίηση (*Dual Scaling*), Αθροιστική Βαθμονόμηση (*Additive Scoring*), Βέλτιστη ή Άριστη Κλιμακοποίηση (*Optimal Scaling*), Ανάλυση Ομοιογένειας (*Homogeneity Analysis*) και άλλα (βλέπε Nishisato, 1980).

Σύμφωνα με τους Van Rijckevorsel και De Leeuw (1988) και Beh (2004), τα πρώτα μαθηματικά αποτελέσματα, με τα οποία μπορούν να συνδεθούν μεθοδολογικά στοιχεία της ΠΑΑ, παρουσιάστηκαν από τον Karl Pearson στην πρώτη δεκαετία του 20^{ου} αιώνα. Η ΠΑΑ, ως μέθοδος ανάλυσης κατηγορικών δεδομένων, εμφανίστηκε στα μέσα της δεκαετίας του '30 με την εργασία του Hirschfeld (1935) και με τις απαρχές της συνδέονται ονόματα και άλλων σημαντικών ερευνητών, όπως των Richardson, Kuder, Horst, Fisher, Maung, Guttman και Burt (βλέπε Nishisato 1980, Greenacre 1984, Gauch 1995). Ο Παπαδημητρίου (2002) αναφέρει ότι αν και η πρώτη προσπάθεια μαθηματικής διατύπωσης της μεθόδου οφείλεται στον H. O. Hartley (ή Hirschfeld) ωστόσο σε μια πιο ολοκληρωμένη μορφή η μέθοδος παρουσιάστηκε από τον Guttman το 1941. Ο Greenacre (1984) παρατηρεί ότι ο Fisher και ο Guttman, σχεδόν ταυτόχρονα και ανεξάρτητα, παρουσίασαν σε θεωρητικό επίπεδο την ίδια ουσιαστικά μέθοδο αλλά σε διαφορετικό πλαίσιο προβληματικής και εφαρμογών: ο Fisher (1940) τη διμεταβλητή εκδοχή της μεθόδου στο χώρο της Βιομετρίας, ενώ ο Guttman (1941) την πολυμεταβλητή εκδοχή στο χώρο της Ψυχομετρίας. Σε κάθε περίπτωση, η μέθοδος δεν παρουσιάστηκε με τη σημερινή της διεθνή ονομασία, *Analyse Factoriel des Correspondances-A.F.C.* στα γαλλικά (*Correspondence Analysis-C.A.* στα αγγλικά), η οποία αποδίδεται στον Benzécri (Nishisato 1980, Van Rijckevorsel & De Leeuw 1988). Ο Benzécri, αφού συστηματοποίησε τη μαθηματική της θεμελίωση ήδη από τη δεκαετία του 1960 (Καραπιστόλης, 1999), την ανήγαγε τελικά σε ένα γενικό σύστημα στατιστικής ανάλυσης δεδομένων (Clausen, 1998).

Αξίζει να σημειωθεί ότι στις αρχές του 1950 παραλλαγές της μεθόδου γνωρίζουν σημαντική ανάπτυξη και διάδοση στην Ιαπωνία, χάρη στις εργασίες του Hayashi (Van Rijckevorsel & De Leeuw 1988, Greenacre & Blasius 1994, Nishisato 1994 και 1980) και σχεδόν ταυτόχρονα στην Αγγλία από τον Burt (1950). Αργότερα, στη δεκαετία του 1970 η μέθοδος γίνεται γνωστή στον Καναδά από το Nishisato με την ονομασία Δυϊκή Κλιμακοποίηση (*Dual Scaling*) (Nishisato 1996, 1994, 1993, 1980 και 1978), ενώ την ίδια περίοδο οι De Leeuw, Young και Takane συστηματοποιούν υπολογιστικά και προγραμματιστικά τις μεθόδους Βέλτιστης Κλιμακοποίηση (*Optimal Scaling*) (De Leeuw, Young & Takane 1976, Young, De Leeuw & Takane 1976, Takane, Young & De Leeuw 1977, Young, Takane & De Leeuw 1978), στο πλαίσιο των οποίων η ΠΑΑ μπορεί να ενταχθεί ως ειδική περίπτωση (Tenenhaus & Young, 1985). Αποτέλεσμα αυτής της προσπάθειας ήταν η δημιουργία και ανάπτυξη της Ολλανδικής Σχολής.

2.1. Η Γαλλική Σχολή της Ανάλυσης Δεδομένων

Στη Γαλλία οι μέθοδοι της ΑΔ γνώρισαν σημαντική ανάπτυξη, ιδιαίτερα μετά το 1970, χάρη στο έργο και την “εμμονή” του Jean-Paul Benzécri Καθηγητή του Πανεπιστημίου Pierre et Marie Curie - Paris VI, ο οποίος συστηματοποίησε τις μαθηματικές βάσεις των μεθόδων, τις ανέδειξε από τη λήθη και κατόρθωσε να αποκτήσουν στο χώρο της Στατιστικής θέση αντάξια της σημαντικότητας και της χρησιμότητάς τους (βλέπε Greenacre 1984, Van Meter *et al.* 1994, Clausen 1998, Καραπιστόλης 1999, Μεϊμάρης 2002, Παπαδημητρίου 2002 και 1994, Le Roux & Rouanet 2004). Ο Αθανασιάδης (1995, σ. 15) γράφει χαρακτηριστικά ότι “Ο Benzécri είναι που θα προσδώσει στο όρο της ανάλυσης δεδομένων μια διάσταση ριζοσπαστική αν όχι διάσταση πολεμικής.” Θεμελίωσε με το έργο του μια Φιλοσοφία και μια

Σχολή την οποία ασπάστηκαν και βελτίωσαν αρκετοί συνεργάτες και μαθητές του όπως οι Diday, Lebart, Escofier, Le Roux, J. Pagès, J.-P. Pagès, Morineau, Tenenhaus, Fenelon, Saporta, Greenacre, Παπαδημητρίου και άλλοι. Η δυναμική των μεθόδων ενισχύθηκε με την έκδοση των επιστημονικών περιοδικών *Cahiers d'Analyse des Données* και *Revue de Statistique Appliquée*, στα οποία δημοσιεύτηκε σημαντικός αριθμός εργασιών τόσο σε θεωρητικό όσο και σε επίπεδο εφαρμογών (Le Roux & Rouanet, 2004). Οι μέθοδοι άρχισαν να διδάσκονται σε πολλά μεταπτυχιακά τμήματα στατιστικής, ενώ σε επίπεδο εφαρμογών η ΠΑΑ, και ιδιαίτερα η πολυμεταβλητή εκδοχή της, έγινε το βασικό στατιστικό εργαλείο για την ανάλυση πολυδιάστατων κατηγορικών δεδομένων που προέρχονταν κυρίως από τη συλλογή ερωτηματολογίων και την ανάλυση κειμένων.

Η Γαλλική παράδοση έχει δώσει ιδιαίτερη έμφαση στη γεωμετρική θεώρηση της ερμηνείας των δεδομένων (Clausen 1998, Le Roux & Rouanet 2004) με βασικό σκοπό την ανάδειξη της ενδογενούς δομής που τα χαρακτηρίζει, η οποία, συνήθως, δεν είναι άμεσα αντιληπτή (Καραπιστόλης, 1999), αλλά βρίσκεται σε λανθάνουσα μορφή. Για τον Benzécri μεγαλύτερη αξία έχουν οι «παράγοντες», δηλαδή οι διαστάσεις και οι δομές που αναδεικνύονται μέσω των μεθόδων και όχι τα ίδια τα δεδομένα, τα οποία αποτελούν μόνο μια προσεγγιστική εικόνα της πραγματικότητας (Van Meter *et al.*, 1994). Ως σχολή έχει να αναδείξει δύο κυρίως οικογένειες μεθόδων (Benzécri & *Collaborateurs*, 1973), την Παραγοντική Ανάλυση των Αντιστοιχιών-ΠΑΑ (*Analyse Factoriel des Correspondances, A.F.C.*) και την Ταξινόμηση κατά Αύξουσα Ιεραρχία (*Classification Ascendante Hierarchique, C.A.H.*)¹. Οι Αναστασιάδου και Παπαδημητρίου (2001, σ. 327) υποστηρίζουν ότι οι μέθοδοι αυτές είναι οι πιο κατάλληλες για την ανάλυση ποιοτικών μεταβλητών και είναι δυνατό να αναδείξουν “*απρόβλεπτες διαστάσεις και να δημιουργήσουν νέες θεωρητικές προσεγγίσεις και προεκτάσεις*” κατά τη διερεύνηση ενός φαινομένου. Ειδικότερα, η ΠΑΑ, έφθασε να κατέχει στη Γαλλία μια τόσο ισχυρή θέση στη μεθοδολογία στατιστικής επεξεργασίας, ώστε να γίνει συνώνυμη με την έννοια της ΑΔ (Clausen, 1998). Μέσω της ΠΑΑ είναι δυνατή η, σχεδόν, καθολική περιγραφή του υπό εξέταση φαινομένου (Αναστασιάδου & Παπαδημητρίου, 2001), το οποίο παρουσιάζεται μέσα από ένα πίνακα κατηγορικών δεδομένων της μορφής «*αντικείμενα × μεταβλητές*» (βλέπε Παπαδημητρίου, 2004 και 1994). Βασικό πλεονέκτημα της μεθόδου είναι η δυνατότητα της ταυτόχρονης γραφικής απεικόνισης των μεταβλητών και των αντικειμένων, δηλαδή της οπτικής αναπαράστασης των αλληλεπιδράσεων και των σχέσεων τους σε ένα κοινό διάγραμμα (Hair *et al.* 1995, Bendixen 1995). Η μόνη απαίτηση της μεθόδου αφορά στις κλίμακες μέτρησης των μεταβλητών, οι οποίες θα πρέπει να είναι ονομαστικές (*nominal*) ή/και διάταξης (*ordinal*). Βέβαια, μπορούν να χρησιμοποιηθούν και ποσοτικές μεταβλητές, αφού πρώτα οι τιμές τους χωριστούν και ομαδοποιηθούν σε κλάσεις με βάση λογικά ή/και στατιστικά κριτήρια (Παπαδημητρίου, 1994). Με τον τρόπο αυτό αναδεικνύονται ιδιότητες (ποιοτικά χαρακτηριστικά) μέσα σε κάθε μεταβλητή και μέσω της ΠΑΑ οπτικοποιούνται και διαπιστώνονται οι μεταξύ τους ομοιότητες ή αντιθέσεις.

Συμπερασματικά, η Γαλλική Σχολή ΑΔ με κύριο εκφραστή τον Benzécri αποτελεί όχι μόνο μια εναλλακτική μεθοδολογική προσέγγιση αλλά και μια νέα φιλοσοφική θεώρηση της Στατιστικής (βλέπε Van Meter *et al.* 1994, Le Roux & Rouanet 2004), μια νέα αντίληψη, ανεξάρτητη από υποδείγματα (μοντέλα), χωρίς *a priori* υποθέσεις και χωρίς αυστηρές πιθανοθεωρητικές προϋποθέσεις, που σπάνια ικανοποιούνται στην πράξη (Gifi, 1996). Κεντρικό ρόλο στο μεθοδολογικό πλαίσιο της Γαλλικής προσέγγισης παίζει η ΠΑΑ, τόσο στη διμεταβλητή όσο και στην πολυμεταβλητή εκδοχή της, για την οποία ο Παπαδημητρίου (2004) τονίζει ότι: “*Η Παραγοντική Ανάλυση των Αντιστοιχιών είναι η σημαντικότερη μέθοδος*

¹ Στη Γαλλία διαδεδομένες είναι επίσης η Συμβολική Ανάλυση (βλέπε Φλώρου, 1997) και η Συνεπαγωγική Στατιστική (βλέπε Gras 1995, Καραπιστόλης 1999 και 1996).

που πρέπει να γνωρίζει κάποιος που επιθυμεί να αναλύσει ένα πολυμεταβλητό φαινόμενο χρησιμοποιώντας την αναλυτική διεξόδου της πολυδιάστατης στατιστικής ανάλυσης.” (σ. 3)

2.2. Η Ολλανδική Σχολή της Ανάλυσης Δεδομένων

Η εισαγωγή στη μεθοδολογία και τη λογική της ΠΑΑ μπορεί να γίνει με πολλούς και διαφορετικούς τρόπους (Benzécri & *Collaborateurs* 1973, Nishisato 1980, Greenacre 1984, Tenenhaus & Young 1985, Israëls 1987, Bekker & De Leeuw 1988, Weller & Romney 1990, Andersen 1991, Benzécri 1992, Gifi 1996). Αυτός είναι, πιθανόν, ο λόγος για τον οποίο η μέθοδος “εφευρέθηκε” αρκετές φορές κατά τη διάρκεια του 20^{ου} αιώνα (Michailidis & De Leeuw, 1998). Αυτή η παράλληλη ανάπτυξη είχε ως αποτέλεσμα να δημιουργηθούν διάφορες προσεγγίσεις και κατευθύνσεις, τόσο σε σχέση με το αλγοριθμικό όσο και σε σχέση με το θεωρητικό υπόβαθρο της μεθόδου.

Σε εργασίες ομάδας ερευνητών του Πανεπιστημίου του Leiden (Department of Data Theory of the Faculty of Social Sciences) έγινε προσπάθεια ενοποίησης των διαφόρων παραλλαγών της μεθόδου αρχικά με την ονομασία *GIFI System* (De Leeuw 1984, Gifi 1996, Michailidis & De Leeuw 1998) και στη συνέχεια ως *Data Theory Scaling System-D.T.S.S.* (Meulman, 1999). Η ενοποίηση έγινε εφικτή με τη συστηματοποίηση και χρήση των μεθόδων της Βέλτιστης Κλιμακοποίησης (Van de Geer 1993α και 1993β, Greenacre 1993, Gifi 1996, De Leeuw 2005α). Σκοπός των μεθόδων αυτών είναι ο μετασχηματισμός ποιοτικών-κατηγορικών μεταβλητών σε ποσοτικές (Young 1981, Van Rijckevorsel & De Leeuw 1988, Greenacre 1993). Αυτό επιτυγχάνεται με την ανάθεση βέλτιστων βαθμών και «βαρών» στις γραμμές (αντικείμενα) και στις στήλες (μεταβλητές) αντίστοιχα του πίνακα δεδομένων (Gifi, 1996). Οι νέες ποσοτικοποιημένες μεταβλητές μπορούν, στη συνέχεια, να χρησιμοποιηθούν σε στατιστικές διαδικασίες όπου απαιτούνται ποσοτικές μεταβλητές (Nishisato 1980, De Leeuw 2005β και 1993, Greenacre 1993, Meulman & Heiser 2004, Le Roux & Rouanet 2004).

Η βελτιστοποίηση είναι έννοια σχετική γιατί επιτυγχάνεται με βάση τα διαθέσιμα δεδομένα που αναλύονται κάθε φορά. Δηλαδή, η κλίμακα μέτρησης μιας μεταβλητής αποκτά νόημα σε σχέση με τις υπόλοιπες. Τα κριτήρια βελτιστοποίησης είναι πολλά, όπως, για παράδειγμα, η βέλτιστη διάκριση μεταξύ των αντικειμένων (δειγματοληπτικές ή πειραματικές μονάδες), η μεγιστοποίηση της ομοιογένειας ή της εσωτερικής συνέπειας μεταξύ των μεταβλητών και η δημιουργία όσο το δυνατό ισχυρότερων γραμμικών σχέσεων μεταξύ ζευγών μεταβλητών (Nishisato 1980, Gifi 1996, Michailidis & De Leeuw 1998, Meulman 1999, Michailidis & De Leeuw 2005). Έτσι, ο χρήστης των μεθόδων θα πρέπει, κάθε φορά, να επιλέγει το κατάλληλο κριτήριο βελτιστοποίησης καθώς και την κλίμακα ποσοτικοποίησης των μεταβλητών, ανάλογα με τους επιδιωκόμενους στόχους της ανάλυσης. Σε κάθε περίπτωση, μια κατάλληλη μη γραμμική «συνάρτηση απώλειας» (*loss function*) βελτιστοποιείται (Bekker & De Leeuw 1988, Gifi 1996, Michailidis & De Leeuw 2000 και 1998) κάτω από συνθήκες που καθορίζονται από το κριτήριο βελτιστοποίησης, τις κλίμακες μέτρησης των μεταβλητών και τους στόχους της εκάστοτε μελέτης.

Υπολογιστικά, η βελτιστοποίηση επιτυγχάνεται κυρίως με την εφαρμογή του αλγόριθμου *Alternating Least Squares* (Εναλλασσόμενα Ελάχιστα Τετράγωνα) (βλέπε Bekker & De Leeuw 1988, Gifi 1996, Michailidis & De Leeuw 1998), σε αντίθεση με την αλγεβρική προσέγγιση του Benzécri. Όμως, στο πλαίσιο της Ολλανδικής προσέγγισης και η ΠΑΑ μπορεί να θεωρηθεί ως μέθοδος βέλτιστης κλιμακοποίησης που έχει στόχο την καλύτερη δυνατή αναπαράσταση των δεδομένων ενός πίνακα συμπτώσεων σε ένα χώρο με λιγότερες διαστάσεις (Bendixen, 1996). Η βελτιστοποίηση έγκειται στο να βρεθούν, με κατάλληλη αλλαγή της κλίμακας μέτρησης, εκείνες οι τιμές των συντεταγμένων των προβολών των

σημείων γραμμών και στηλών πάνω στους παραγοντικούς άξονες (Weller & Romney, 1990) που μεγιστοποιούν τη διακύμανση κατά τη διεύθυνση των αξόνων (Greenacre, 1993).

Το σύστημα *GIFI* αποτελεί ένα σύνολο μεθόδων για την ανάλυση κυρίως κατηγορικών μεταβλητών και χαρακτηρίζει την Ολλανδική Σχολή ΑΔ (Bond & Michailidis, 1996). Σημαντικός σταθμός στην εξέλιξη των μεθόδων της Σχολής αυτής ήταν η ανάπτυξη της Ανάλυσης Ομοιογένειας (*Homogeneity Analysis*) ή αλλιώς της Πολλαπλής Ανάλυσης των Αντιστοιχιών (*Multiple Correspondence Analysis*) (Michailidis & De Leeuw, 1998), η οποία αποτελεί, ως ένα βαθμό, την πολυμεταβλητή εκδοχή της ΠΑΑ της Γαλλικής Σχολής. Με βάση το θεωρητικό και αλγοριθμικό πλαίσιο αυτής της μεθόδου αναπτύχθηκε η Μη Γραμμική Ανάλυση σε Κύριες Συνιστώσες (*Non Linear Principal Component Analysis* ή *Principal Components Analysis for Categorical Data*) και η Μη Γραμμική Κανονικοποιημένη Συσχέτιση (*Non Linear Canonical Correlation*). Η ερευνητική δραστηριότητα των Ολλανδών επεκτάθηκε και στην ανάπτυξη μεθόδων με τις οποίες είναι δυνατή η ταυτόχρονη χρήση δεδομένων μικτού τύπου (κατηγορικά και ποσοτικά) καθώς και στην δυνατότητα ανάλυσης ομάδων μεταβλητών (Van der Burg & De Leeuw 1988, Van der Burg, De Leeuw & Dijksterhuis 1994, Michailidis & De Leeuw 1998). Οι αντίστοιχες στατιστικές διαδικασίες περιλαμβάνονται στο υποσύστημα *Categories* του στατιστικού πακέτου SPSS (Meulman & Heiser, 2004).

Το όνομα *GIFI* αντιπροσωπεύει και την ίδια την ομάδα των Ολλανδών ερευνητών που για πρακτικούς κυρίως λόγους το χρησιμοποίησαν για κοινές δημοσιεύσεις. Μέλη της ομάδας αποτέλεσαν οι Bettonvil, Van der Burg, Van de Geer, Heiser, Meulman, Van Rijckevorsell, Stoor και άλλοι (Gifi, 1996) με πρωταγωνιστικό και καθοδηγητικό ρόλο αυτόν του Jan de Leeuw (Greenacre, 1993). Σημαντικό μέρος του ερευνητικού έργου των Ολλανδών ερευνητών έχει δημοσιευθεί σε τεχνικές αναφορές του Πανεπιστημίου του Leiden και στο επιστημονικό περιοδικό *Psychometrika* (βλέπε Van Rijckevorsel & De Leeuw, 1988).

Το ενδιαφέρον της Ολλανδικής Σχολής εστιάζεται στην ανάδειξη των μη γραμμικών σχέσεων μεταξύ των μεταβλητών που προκύπτουν λόγω της κατηγορικής φύσης των δεδομένων (Van der Burg & De Leeuw 1983, Bekker & De Leeuw 1988, Van de Geer 1993α και 1993β, Heiser & Meulman 1994, Gifi 1996, Michailidis & De Leeuw 1998). Η προσέγγιση αυτή έρχεται σε αντίθεση με παραδοσιακές μεθόδους, όπως για παράδειγμα η Πολλαπλή Γραμμική Παλινδρόμηση, η Ανάλυση Διακύμανσης και η Ανάλυση σε Κύριες Συνιστώσες, οι οποίες προϋποθέτουν γραμμικές σχέσεις και υποδείγματα (Gifi, 1996).

Συνοψίζοντας, οι μέθοδοι της Ολλανδικής Σχολής αποτελούν μια επέκταση των μεθόδων της Γαλλικής Σχολής ΑΔ. Οι δύο σχολές έχουν αρκετές ομοιότητες αλλά και διαφορές. Έτσι, ενώ η ΠΑΑ στην πολυμεταβλητή της εκδοχή αποτελεί μία από τις σημαντικότερες μεθόδους και για τις δύο σχολές, ωστόσο το θεωρητικό και μεθοδολογικό πλαίσιο στο οποίο αναπτύχθηκε η μέθοδος σε κάθε σχολή είναι διαφορετικό. Και στις δύο σχολές, ισχυρό κίνητρο εξέλιξης και διάδοσης των προτεινόμενων μεθόδων αποτελεί ο διαγραμματικός τρόπος παρουσίασης και διάχυσης της παραγόμενης πληροφορίας. Υπάρχει όμως σημαντική διαφορά σε ό,τι αφορά την προβληματική, τη θεωρητική και την υπολογιστική προσέγγιση που ακολουθεί η κάθε Σχολή. Οι μέθοδοι της Ολλανδικής Σχολής αντιμετωπίζονται τελικά ως προβλήματα μη γραμμικής βελτιστοποίησης υπό συνθήκες. Η επίλυση τους επιτυγχάνεται με τη χρήση επαναληπτικών υπολογιστικών αλγορίθμων και όχι αλγεβρικά. Τέλος, η ιδιαίτερη προσήλωση των Ολλανδών ερευνητών στην ανάδειξη μη γραμμικών σχέσεων στα δεδομένα είχε ως αποτέλεσμα οι μέθοδοι που αναπτύχθηκαν να χαρακτηρίσουν ένα νέο πεδίο έρευνας της Στατιστικής, αυτό που οι ίδιοι ονόμασαν *Μη Γραμμική Πολυμεταβλητή Ανάλυση* (Bekker & De Leeuw 1988, Gifi 1996).

3. Η Ανάλυση Δεδομένων και η Στατιστική

3.1. Οι Απόψεις του Tukey

Ο John Tukey, διακεκριμένος ερευνητής στο χώρο της Μαθηματικής Στατιστικής (Cohen & Cohen, 1983), ήταν ο πρώτος που διαχώρισε την ΑΔ από τη Στατιστική και την παρουσίασε ως ανεξάρτητη επιστήμη (Gifi, 1996). Κατά την άποψή του (Tukey, 1962) η ΑΔ αποτελεί ένα γενικότερο μεθοδολογικό πλαίσιο απ' ό,τι η Επαγωγική Στατιστική, ενώ υπάρχουν πεδία της Μαθηματικής Στατιστικής που είναι έξω από το πλαίσιο της ΑΔ. Σύμφωνα με τον Tukey, η ΑΔ περιλαμβάνει: α) μεθόδους επεξεργασίας και ανάλυσης των δεδομένων, β) τεχνικές ερμηνείας των αποτελεσμάτων, γ) μεθοδολογίες σχεδιασμού των διαδικασιών συλλογής των δεδομένων και δ) όλα τα μαθηματικά και στατιστικά “εργαλεία – βοηθήματα” που μπορούν να χρησιμοποιηθούν με βασικό σκοπό την ευκολότερη ανάλυση και ερμηνεία των δεδομένων. Για τον Tukey, στην ΑΔ θα πρέπει να δοθεί ιδιαίτερα μεγάλη έμφαση στην κρίση. Τουλάχιστον τρία διαφορετικά είδη ή πηγές κρίσης είναι δυνατό να εμπλέκονται σχεδόν σε κάθε περίπτωση (Tukey, 1962)

“(α1) κρίση που βασίζεται σε εμπειρία από το συγκεκριμένο θεματικό πεδίο από το οποίο προέρχονται τα δεδομένα

(α2) κρίση που βασίζεται πάνω σε μία ευρεία εμπειρία για τον τρόπο που έχουν λειτουργήσει διάφορες τεχνικές ανάλυσης δεδομένων σε σχέση με ποικίλα πεδία εφαρμογής

(α3) κρίση που βασίζεται πάνω σε αφηρημένα αποτελέσματα για τις ιδιότητες συγκεκριμένων τεχνικών, που έχουν ληφθεί είτε μέσω μαθηματικών αποδείξεων είτε μέσω εμπειρικής δειγματοληψίας.” (σ. 9).

Η συνεισφορά του Tukey στην ΑΔ αφορά σε ένα πλήθος ευέλικτων γραφικών κυρίως μεθόδων και τεχνικών για τη διερευνητική και περιγραφική στατιστική ανάλυση των δεδομένων (βλέπε Tukey 1977, Hoaglin 2003). Ο Tukey θεωρεί ότι η προσέγγιση αυτή συνιστά κατά βάση μια Φιλοσοφία, η οποία επαναπροσανατολίζει τη Στατιστική στους αρχικούς της στόχους, όπου η στατιστική περιγραφή των δεδομένων παίζει το σημαντικότερο ρόλο. Δεν αντιτίθεται στον επιβεβαιωτικό χαρακτήρα της Επαγωγικής Στατιστικής και διατυπώνει τελικά την άποψη ότι και οι δύο προσεγγίσεις (Επαγωγική Στατιστική και Ανάλυση Δεδομένων), μπορούν και σε ορισμένες περιπτώσεις επιβάλλεται, να χρησιμοποιούνται συμπληρωματικά (Tukey, 1980 και 1977). Σε κάθε περίπτωση, η διερευνητική – περιγραφική προσέγγιση είναι αυτή που θα πρέπει να προηγείται γιατί, σύμφωνα με την άποψή του (1980), μερικές φορές είναι πιο δύσκολο αλλά και πιο ενδιαφέρον να διατυπώνεις ερωτήσεις από το να παίρνεις απαντήσεις.

3.2. Οι Αρχές του Benzécri

Οι αρχές της Γαλλικής Σχολής ΑΔ, όπως εξηγούνται από τον Benzécri, διαφέρουν αρκετά από αυτές του Tukey. Ο Benzécri διατυπώνει πέντε βασικές αρχές που ορίζουν την ΑΔ (Benzécri & Collaborateurs, 1973):

“Αρχή 1: Η Στατιστική και η Θεωρία Πιθανοτήτων δεν είναι το ίδιο. Αρκετοί συγγραφείς (οι οποίοι, και σας το λέω στα Γαλλικά, σπάνια γράφουν στη γλώσσα μας...) έχουν δομήσει μία πομπώδη επιστήμη υπό την ονομασία «Μαθηματική Στατιστική», η οποία βρίθεται υποθέσεων που ποτέ δεν ικανοποιούνται στην πράξη. Δεν μπορούμε να προσδοκούμε μία λύση στα τυπολογικά μας προβλήματα από αυτούς τους συγγραφείς. (σ. 3).

Αρχή 2: Το μοντέλο θα πρέπει να ακολουθεί τα δεδομένα, και όχι το αντίστροφο. Αυτό είναι ένα ακόμα λάθος στην εφαρμογή των μαθηματικών στις Ανθρωπιστικές Επιστήμες: η αφθονία δηλαδή των μοντέλων, που δομούνται a priori και μετά έρχονται αντιμέτωπα με τα δεδομένα μέσω του επονομαζόμενου «ελέγχου». Συχνά ο «έλεγχος» χρησιμοποιείται για να δικαιολογήσει

κάποιο μοντέλο, στο οποίο ο αριθμός των παραμέτρων που πρέπει να ενταχθούν είναι μεγαλύτερος από τον αριθμό των δεδομένων σημείων. Και συχνά χρησιμοποιείται, αντιθέτως, για να απορρίψει κατηγορηματικά ως αβάσιμα, ακόμα και τα πιο επικριτικά σχόλια του πειραματιστή. Αυτό που χρειαζόμαστε όμως είναι μία αυστηρή μέθοδος για να εξάγουμε τη δομή, ξεκινώντας από τα δεδομένα. (σ. 6).

Αρχή 3: Είναι βολικό να χειρίζεται κανείς ταυτόχρονα πληροφορίες σε όσες περισσότερες διαστάσεις μπορεί. Ως συνέπεια, το πρόβλημα της εγκυρότητας ενός «ελέγχου» - το οποίο, ομολογουμένως, είναι ορισμένες φορές δύσκολο - δεν παρουσιάζεται πλέον ως τόσο σημαντικό. Κανείς δεν γνωρίζει αν η ανισότητα $0,5 \neq 0,7$ θα πρέπει να ερμηνευτεί σε αυτές τις πρακτικές περιπτώσεις ως ένα συγκεκριμένο εμπειρικό αποτέλεσμα, ή απλώς ως τυχαίο αποτέλεσμα. Όμως το να βρει κανείς ότι σε ένα χώρο δύο διαστάσεων υπάρχουν πενήντα σημεία κατά προσέγγιση τοποθετημένα σε έναν κύκλο είναι πραγματικά μία ανακάλυψη (τουλάχιστον αν η μέθοδος υπολογισμού δεν μας παραπλανά!). (σ. 9).

Αρχή 4: Για την ανάλυση περίπλοκων στοιχείων, και ειδικά για την ανάλυση κοινωνικών δεδομένων, χρειαζόμαστε απαραίτητως τον ηλεκτρονικό υπολογιστή. Η αρχή αυτή είναι προφανώς ορθή ... αλλά τι άποψη θα είχαν γι' αυτό οι φοβεροί και τρομεροί πατέρες μας δεκαπέντε χρόνια πριν; (σ. 12).

Αρχή 5: Η χρήση του υπολογιστή σημαίνει ότι όλες οι τεχνικές που σχεδιάστηκαν πριν την έλευση του αυτόματου υπολογισμού θα πρέπει να εγκαταλειφθούν. Και λέω τεχνικές και όχι επιστήμη: οι γεωμετρικές και αλγεβρικές αρχές των προγραμμάτων μας ήταν γνωστές στον Laplace, πριν 150 χρόνια. Όμως ο Laplace ήταν και ο συγγραφέας μίας πραγματείας πάνω στη ουράνια μηχανική, η οποία μόλις τώρα επανεκδόθηκε για να χρησιμοποιηθεί από μηχανικούς του διαστήματος ... Και να φανταστεί κανείς ότι η πραγματεία αυτή δεν ήταν αρκετή για να κατακτήσει ο Ναπολέων το φεγγάρι.” (σ. 15).

Οι Ολλανδοί ερευνητές δεν συμφωνούν απόλυτα με τον αυστηρό και απόλυτο τρόπο που διατυπώνει ο Benzécri τις πέντε αρχές για το τι είναι η ΑΔ (Gifi, 1996). Για παράδειγμα, δεν συμφωνούν με αυτό που ο Benzécri φαίνεται να υπονοεί στις Αρχές 3, 4 και 5 ότι δηλαδή οι μεγάλες, αδόμητες πολυμεταβλητές ομάδες δεδομένων είναι η μόνη δυνατή κατάσταση έρευνας. Εν γένει συμφωνούν με τις Αρχές 2 και 3, αλλά επισημαίνουν ότι αν οι αρχές αυτές εφαρμοστούν συστηματικά είναι δυνατό να οδηγήσουν σε ένα “τυφλό” εμπειρισμό, ο οποίος τρομοκρατεί εν γένει τους Ψυχομέτρους. Πιστεύουν ότι τουλάχιστον κάποια θεωρία – μοντέλο θα πρέπει να καθοδηγεί τον χρήστη των μεθόδων, για παράδειγμα στην επιλογή των σχετικών με το υπό εξέταση φαινόμενο μεταβλητών, ώστε να είναι εφικτή η ερμηνεία των αποτελεσμάτων. Για τους Ολλανδούς η Αρχή 1 του Benzécri είναι ίσως η πιο σημαντική. Τα πιο ενδιαφέροντα σημεία διατυπώνονται από τον ίδιο τον Benzécri (Benzécri & Collaborateurs, 1973):

“Τα μαθηματικά θεμέλια της στατιστικής ανάλυσης είναι περισσότερο αλγεβρικά και γεωμετρικά (και όταν εμπλέκονται πολλές διαστάσεις, οι γεωμετρικές ιδέες συγχωνεύονται με τους αλγεβρικούς υπολογισμούς) και μετά πιθανολογικά. Είναι προτιμότερο να μιλάει κανείς για κύριους άξονες κ.τ.λ. ... που καθορίζονται σε σχέση με έναν πεπερασμένο αριθμό πραγματικών δεδομένων, από το να μιλάει για τη μέση τιμή, κ.τ.λ. ... που ορίζεται σε ένα δυναμικά άπειρο Δειγματοχώρο. Όμως οι πιθανολογικές έννοιες και ιδέες μπορούν να αναδείξουν την αναγκαιότητα αλγεβρικών πράξεων και μερικές φορές μπορούν να χρησιμοποιηθούν για να αξιολογήσουν την χρησιμότητά τους.” (σ. 6).

Για τον Benzécri η ΑΔ αποτελεί μια διαδικασία μάθησης και απόκτησης γνώσης μέσα από ποιοτικές και ποσοτικές καταγραφές πραγματικών εμπειριών, κατά τις οποίες η φροντίδα και η ευθύνη της ερμηνείας των αποτελεσμάτων και των συνεπειών τους αφήνεται κυρίως στους ερευνητές – χρήστες των μεθόδων και όχι σε πιθανολογικούς μηχανισμούς.

3.3. Οι Θέσεις των Ολλανδών

Στο πλαίσιο της Ολλανδικής Σχολής η ΑΔ είναι συνώνυμη με τη Στατιστική (De Leeuw, 2005β), ενώ η Μαθηματική Στατιστική είναι ενδιαφέρουσα στο βαθμό που τα πορίσματά της έχουν πρακτικές εφαρμογές (Gifi, 1996) ιδιαίτερα στη μελέτη και στον έλεγχο της “σταθερότητας” των αποτελεσμάτων που παράγονται από τις μεθόδους ΑΔ (Markus, 1994α και 1994β). Ο όρος “σταθερότητα” αναφέρεται στο πόσο συνεπή είναι τα αποτελέσματα κάτω από διαταραχές κάποιων αρχικών συνθηκών, οι οποίες είτε αφορούν στη μεθοδολογία συλλογής των δεδομένων είτε στις ίδιες τις μεθόδους ανάλυσης, όπως για παράδειγμα στους εκάστοτε αλγόριθμους καθώς και στους αριθμητικούς υπολογισμούς. Πιο συγκεκριμένα, φαίνεται να ανοίγεται ένας μεγάλος ορίζοντας σε ό,τι αφορά τη μελέτη και τον έλεγχο της σταθερότητας (Gifi 1996, Michailidis 1996, Michailidis & De Leeuw 1998): α) των αποτελεσμάτων, στη θεωρητική περίπτωση επανάληψης της έρευνας κάτω από τις ίδιες συνθήκες (*replication stability*), β) των διαφόρων δεικτών (αναφοράς, ερμηνείας και ποιότητας) με βάση τους οποίους γίνεται ή/και διευκολύνεται η παρουσίαση και η ερμηνεία των αποτελεσμάτων (*statistical stability*), γ) των αποτελεσμάτων κάτω από διαφορετικές μεθόδους συλλογής των δεδομένων ή σύνθεσης (απαλοιφή αντικειμένων ή/και μεταβλητών) του πίνακα δεδομένων που θα αναλυθεί (*stability under data selection*), δ) των αποτελεσμάτων κάτω από διαφορετικά υποδείγματα διερεύνησης συσχετίσεων ή σχέσεων αιτίας-αποτελέσματος (*stability under model selection*), ε) της ακρίβειας των αριθμητικών υπολογισμών είτε λόγω σφαλμάτων στρογγυλοποίησης είτε λόγω των επαναληπτικών υπολογιστικών μεθόδων που συνήθως χρησιμοποιούνται (*numerical stability*), στ) των αποτελεσμάτων κάτω από διαταραχές των αρχικών δεδομένων και κάτω από την ισχύ Κεντρικών Οριακών Θεωρημάτων της Θεωρίας Πιθανοτήτων (*analytical and algebraic stability*) και ζ) των αποτελεσμάτων κάτω από την επίδραση που ενδεχομένως να έχει σε αυτά η επιλογή μιας διαφορετικής μεθόδου ανάλυσης (*stability under selection of technique*). Ένα άλλο χαρακτηριστικό της Ολλανδικής Σχολής είναι η μετάθεση του ενδιαφέροντος από τις ιδιότητες και τις εφαρμογές της Πολυδιάστατης Κανονικής και της Πολυωνυμικής Κατανομής (Gifi, 1996) σε μια ενδιάμεση κατάσταση με τη χρήση μεθόδων επαναδειγματοληψίας (βλέπε Efron & Tibshirani 1993, DiCiccio & Efron 1996). Τεχνικές όπως η *Monte Carlo* και η *Bootstrap* κατέχουν σημαντική θέση κυρίως στους ελέγχους εξωτερικής και εσωτερικής εγκυρότητας των παραγόμενων αποτελεσμάτων (βλέπε Lebart *et al.* 1984, Greenacre 1993 και 1984, Markus 1994α και 1994β, Gifi 1996, Michailidis 1996, Chateau & Lebart 1996, Μπεχράκης 1999, Lebart 2005).

Έτσι, για τους Ολλανδούς η Θεωρία Πιθανοτήτων, η Στατιστική, η Υπολογιστική Στατιστική, τα Μαθηματικά, τα Υπολογιστικά Μαθηματικά και η Πληροφορική όχι μόνο ενσωματώνονται στην ΑΔ αλλά μπορούν να συνδυαστούν με αυτήν, με τρόπο ώστε να δημιουργηθούν νέοι τομείς έρευνας κυρίως σε σχέση με τη σταθερότητα των αποτελεσμάτων, η οποία εγγυάται, τελικά, την αξιοπιστία και την εγκυρότητα των συμπερασμάτων.

4. Συζήτηση

Είναι γεγονός ότι συχνά δεν γνωρίζουμε τους νόμους στους οποίους ένα φαινόμενο υπακούει ή το μαθηματικό υπόδειγμα (ντετερμινιστικό ή στοχαστικό), το οποίο καθορίζει τη συμπεριφορά του, είτε γιατί είναι αρκετά σύνθετο και δεν επιδέχεται λεπτομερή μαθηματική περιγραφή είτε δεν μπορεί να διατυπωθεί κάποιος νόμος που να εκφράζει μια αποδεκτή σχέση αιτίας-αποτελέσματος (Δερμάνης, 1986). Στην περίπτωση αυτή, αναδεικνύεται η χρησιμότητα της ΑΔ, η οποία ως κλάδος της Πολυδιάστατης Στατιστικής Ανάλυσης είναι δυνατόν να περιγράψει την πολυπλοκότητα ορισμένων φαινομένων και να αναδείξει τυχόν “άγνωστες – μη εμφανείς” πτυχές ή ιδιαιτερότητες της συμπεριφοράς τους.

Ο χαρακτηρισμός των μεθόδων της ΑΔ ως στατιστικές διαδικασίες ανεξάρτητες από στοχαστικά υποδείγματα, οι οποίες χρησιμοποιούνται χωρίς *a priori* υποθέσεις και προϋποθέσεις, είναι ορθός μόνο σε μια πρώτη θεώρηση και πάντα σε σύγκριση με τις διαδικασίες της Επαγωγικής Στατιστικής, οι οποίες έχουν αρκετές πιθανοθεωρητικές προϋποθέσεις και εφαρμόζονται για την απόρριψη ή όχι συγκεκριμένων ερευνητικών υποθέσεων. Οι προϋποθέσεις αυτές δεν συνδέονται με τις γενικότερες επιστημονικές υποθέσεις, οι οποίες προηγούνται και αποτελούν το κίνητρο για την πραγματοποίηση μιας έρευνας, αλλά είναι απαιτήσεις μόνο των στατιστικών τεχνικών (Μπεχράκης, 1999). Η απόλυτη αποδοχή της παραπάνω θέσης θα μπορούσε να οδηγήσει σε μια υπέρμετρη απλοποίηση της στατιστικής σκέψης όπου το ενδιαφέρον θα επικεντρώνεται μόνο στην εξεύρεση ορθών, κατά περίπτωση, λύσεων θέτοντας στο περιθώριο τη μαθηματική αυστηρότητα και τη δυνατότητα γενίκευσης.

Η σημαντικότερη διαφορά των μεθόδων της ΑΔ σε σχέση με αυτές της Επαγωγικής Στατιστικής είναι ότι για την εφαρμογή τους δεν απαιτείται η προσαρμογή των δεδομένων σε κάποιο στοχαστικό υπόδειγμα και η αντίστοιχη συμπερασματολογία δεν υπάγεται σε κάποιο μηχανισμό επαγωγικού συλλογισμού με την έννοια της στατιστικής σημαντικότητας. Η αλήθεια είναι ότι οι μέθοδοι, συνήθως, δεν συνοδεύονται από στατιστικούς ελέγχους. Αυτό όμως δεν σημαίνει ότι δεν είναι δυνατό να υπάρξουν τέτοιοι και μάλιστα με ελάχιστες προϋποθέσεις (Μενεξές, 2007). Επίσης, δεν δίνεται ιδιαίτερη έμφαση στο μηχανισμό συλλογής και συγκρότησης των διαθέσιμων δεδομένων (μέθοδος δειγματοληψίας, πειραματικός σχεδιασμός, ανεξαρτησία των παρατηρήσεων) αρκεί τα δεδομένα να μπορούν να πινακοποιηθούν σε μορφή κατάλληλη για την εφαρμογή της εκάστοτε μεθόδου. Τα δεδομένα, έστω και αν προέρχονται από δείγμα, αντιμετωπίζονται σαν να αποτελούν ολόκληρο τον υπό εξέταση πληθυσμό (Greenacre, 1984) δίνοντας ένα περιγραφικό και διερευνητικό χαρακτήρα στις μεθόδους αποδυναμώνοντας οποιεσδήποτε γενικεύσεις, όπως αυτές νοούνται στο πλαίσιο της Επαγωγικής Στατιστικής.

Κάθε μέθοδος της ΑΔ στηρίζεται σε ένα σύστημα μαθηματικών υποδειγμάτων, τα οποία αποσκοπούν στην “προσομοίωση” του αντίστοιχου φαινομένου. Γενικά, η προσομοίωση είναι εφικτή μέσω ενός συνόλου συμβόλων και εξισώσεων που έχουν ως σκοπό να περιγράψουν τη συμπεριφορά του σε σχέση με συγκεκριμένα χαρακτηριστικά ή ιδιότητες που έχουν επιλεγεί και θεωρούνται σημαντικές (Hillier & Lieberman, 1995). Οι συσχετισμοί μεταξύ των πραγματικών δεδομένων ή αλλιώς των πρωτοτύπων (*prototypes*), που χαρακτηρίζουν το φαινόμενο, αντικαθίστανται από ανάλογους συσχετισμούς μεταξύ μαθηματικών οντοτήτων (*entities*) (Davis & Hersh, 1980). Η βασική απαίτηση είναι το μαθηματικό υπόδειγμα και το φαινόμενο να είναι ισόμορφα σε όλα τα θεμελιώδη, για το πρόβλημα που εξετάζεται, θέματα (Βασιλείου, 1985). Η επιτυχία σε μια τέτοια προσπάθεια έγκειται στην επιλογή και στην κατασκευή ενός υποδείγματος που να περιλαμβάνει όσο το δυνατόν περισσότερη πληροφορία σε σχέση με την πραγματικότητα, όπως αυτή αποτυπώνεται με βάση τα διαθέσιμα δεδομένα. Έτσι, αυτόματα εισάγονται αρχικές προϋποθέσεις και υποθέσεις που αφορούν στον τρόπο με τον οποίο θα μετρηθεί η ποσότητα και η ποιότητα της πληροφορίας, στην καταλληλότητα και αντιπροσωπευτικότητα των δεδομένων, στην κωδικοποίησή τους, στους θεωρούμενους συσχετισμούς μεταξύ των πρωτοτύπων και των αντίστοιχων μαθηματικών οντοτήτων, στις μαθηματικές ιδιότητες των υποδειγμάτων, στη δυνατότητα πραγματοποίησης των διαφόρων υπολογισμών καθώς και στην ακρίβειά τους. Με άλλα λόγια, εισάγονται ορισμένες αρχικές συνθήκες - προϋποθέσεις ως προς την εγκυρότητα και αξιοπιστία των μεθόδων, οι οποίες θα επιλεγούν τελικά ως οι καταλληλότερες για την ανάλυση των δεδομένων. Οι χρήστες των μεθόδων, ανάλογα με το επιστημονικό πεδίο που δραστηριοποιούνται, θα πρέπει κάθε φορά να αξιολογούν τις προϋποθέσεις εγκυρότητας και να οριοθετούν την ισχύ του αντίστοιχου υποδείγματος. Οι αποφάσεις τους είναι δυνατό να επηρεάσουν την παραγόμενη πληροφορία η οποία θα πρέπει

να αποκτήσει νόημα σε ένα συγκεκριμένο ερευνητικό και επιστημονικό πλαίσιο. Συνεπώς, είναι απαραίτητη η θέσπιση κάποιων λογικών κανόνων ή κριτηρίων τουλάχιστον για την επιλογή των μεταβλητών μέσω των οποίων θα περιγραφεί το υπό εξέταση φαινόμενο. Αν υπάρχει κάποια δομή στα δεδομένα οι μέθοδοι θα την αναδείξουν. Η δομή θα γίνει ενδεχομένως πιο ξεκάθαρη αν εισαχθούν στις αναλύσεις νέες μεταβλητές από το ίδιο ερευνητικό γνωστικό πεδίο. Η δομή μάλλον θα εξαφανιστεί αν προστεθούν αρκετές μεταβλητές από άλλα πεδία οι οποίες θα λειτουργήσουν συσκοτιστικά ως “θόρυβος”. Παρόλα αυτά, θα πρέπει να γίνει σαφές ότι ένα υπόδειγμα ή σύστημα μεθόδων ανάλυσης της εμπειρίας δεν μπορεί να χρησιμοποιηθεί για να αποδείξει οτιδήποτε σχετικά με τον πραγματικό κόσμο (De Leeuw, 2005β) αν και μπορεί, σε μερικές περιπτώσεις, να μας βοηθήσει να ανακαλύψουμε κάποια σημαντικά στοιχεία γι’ αυτόν. Ένα υπόδειγμα δεν μπορεί να χαρακτηριστεί σωστό ή λάθος παρά ως χρήσιμο ή μη χρήσιμο (Pfeiffer, 1978). Άλλωστε, οι μέθοδοι της ΑΔ δεν διαθέτουν κάποιο “μαγικό” μηχανισμό, ο οποίος θα λειτουργήσει με τέτοιο τρόπο ώστε πάντα ένας άμορφος και αδόμητος σωρός από δεδομένα θα μετατραπεί σε επιστημονική γνώση.

Σύμφωνα με τον Breiman (2001), ιδιαίτερη έμφαση θα πρέπει να δίνεται στην κατανόηση του εκάστοτε προβλήματος και των αντίστοιχων εμπειρικών δεδομένων πριν την προσαρμογή και τον έλεγχο οποιουδήποτε υποδείγματος. Οι μέθοδοι της ΑΔ μπορούν να συμβάλλουν προς την κατεύθυνση αυτή αποκαλύπτοντας πληροφορία σχετικά με τον συχνά άγνωστο μηχανισμό παραγωγής των δεδομένων. Τέλος, η ΑΔ, ως υποσύνολο των πολυμεταβλητών στατιστικών μεθόδων, έχει να προσφέρει ορισμένα πλεονεκτήματα, τα οποία μπορούν να συμβάλλουν, σε ορισμένες περιπτώσεις, στην επιτυχημένη διεξαγωγή πολλών φάσεων ενός ερευνητικού έργου (Μενεξές, 2008 και 2007):

- Ο μη παραμετρικός χαρακτήρας των μεθόδων, οι οποίες έχουν ελάχιστες τεχνικές προϋποθέσεις και δεν απαιτούν τα διαθέσιμα δεδομένα να χαρακτηρίζονται από συγκεκριμένες στατιστικές ιδιότητες. Το γεγονός αυτό μπορεί να διευκολύνει τη διαδικασία συλλογής και επεξεργασίας των δεδομένων, περιορίζοντας έτσι τον απαιτούμενο χρόνο και το κόστος διεξαγωγής της έρευνας.

- Η δυνατότητα των μεθόδων να χειριστούν και ποιοτικά χαρακτηριστικά (μεταβλητές), τα οποία συχνά δεν διαθέτουν τις απαραίτητες στατιστικές ιδιότητες που προαπαιτούνται σε άλλες στατιστικές προσεγγίσεις.

- Η μη γραμμικότητα των σχέσεων μεταξύ των μεταβλητών. Σε αντίθεση με τις στατιστικές διαδικασίες, οι οποίες στηρίζονται αποκλειστικά σε γραμμικά υποδείγματα, οι μέθοδοι μπορούν να αναδείξουν είτε γραμμικές είτε μη γραμμικές σχέσεις ανάλογα με τα δεδομένα. Τα παραγόμενα στατιστικά αποτελέσματα (δείκτες, διαγράμματα) έχουν απτή φυσική ερμηνεία και είναι δυνατό να αναδείξουν ισχυρούς, αλλά λανθάνοντες, μηχανισμούς που οδήγησαν στη συγκρότηση των δεδομένων, στοιχείο το οποίο θεωρείται απαραίτητο για την ερμηνεία και κατανόηση του υπό εξέταση φαινομένου.

- Η αυξημένη ευελιξία στην ανάπτυξη νέων μεθόδων ή στη βελτίωση των υπαρχόντων με τη χρήση τεχνικών γραμμικής ή μη γραμμικής βελτιστοποίησης, με σκοπό την καλύτερη ερμηνεία των δεδομένων και τον περιορισμό της πολυπλοκότητάς τους.

- Η εφαρμογή τους σε ειδικές περιπτώσεις, όπου άλλες στατιστικές μέθοδοι απλά δεν μπορούν να εφαρμοστούν. Χαρακτηριστικά αναφέρουμε τις παρακάτω καταστάσεις όπου: α) το δείγμα που έχει συγκεντρωθεί είναι συμπτωματικό (μη πιθανότητας) και όχι τυχαίο και β) οι πειραματικοί σχεδιασμοί είναι ατελείς και δεν υπακούν σε δομημένη πορεία (*quasi experimental designs*) (βλέπε Mertens, 1998). Και στις δύο περιπτώσεις δεν μπορούν να εφαρμοστούν, για παράδειγμα, μέθοδοι της Επαγωγικής Στατιστικής.

Τα πλεονεκτήματα που αναφέρθηκαν παραπάνω αν και έχουν, σε πρακτικές εφαρμογές, υψηλή αποτελεσματικότητα, ωστόσο δεν αποτελούν πανάκεια. Δεδομένης της πληθώρας των διαθέσιμων πολυμεταβλητών στατιστικών τεχνικών, η επιλογή της πλέον αποτελεσματικής

μεθοδολογίας, όποια πλεονεκτήματα κι αν έχει, μπορεί να στοιχειοθετηθεί μόνο σε συγκεκριμένα πεδία εφαρμογών και με συγκεκριμένα δεδομένα.

5. Βιβλιογραφία

- Αθανασιάδης, Η. (1995). Παραγοντική Ανάλυση Αντιστοιχιών και Ιεραρχική Ταξινόμηση. Αθήνα: Εκδόσεις ΝΕΩΝ ΤΕΧΝΟΛΟΓΙΩΝ.
- Αναστασιάδου, Σ. & Παπαδημητρίου, Γ. (2001). Χρήση Μεθόδων της Πολυδιάστατης Στατιστικής Ανάλυσης για τον Προσδιορισμό των Διαθέσεων των Φοιτητών προς τη Στατιστική. *Πρακτικά, Τέταρτο Παγκύπριο Συνέδριο Μαθηματικής Παιδείας και Συμπόσιο Αστροναυτικής και Διαστήματος*, (σ.σ. 327-335). Λάρνακα.
- Andersen, E. (1991). *The Statistical Analysis of Categorical Data*. Berlin-Heidelberg: Springer-Verlag.
- Βασιλείου, Π. Χ. (1985). *Στοχαστικές Μέθοδοι στις Επιχειρησιακές Έρευνες*. Θεσσαλονίκη: Εκδόσεις Χριστοδουλίδη.
- Beh, E. (2004). Simple Correspondence Analysis: A Bibliographic Review. *International Statistical Review*, **72**(2), 257–284.
- Bekker, P. & De Leeuw, J. (1988). Relations Between Variants of Non-Linear Principal Components Analysis. In J. Van Rijckevorsel and J. De Leeuw (Eds), *Component and Correspondence Analysis. Dimension Reduction by Functional Approximation*, (pp. 1-31). Chichester: John Willey & Sons Ltd.
- Bendixen, M. (1995). Compositional Perceptual Mapping Using Chi-Squared Trees Analysis and Correspondence Analysis. *Journal of Marketing Management*, **11**, 571-581.
- Benzécri J.-P. & Collaborateurs (1973). *L'Analyse des Données. Vol. 1: Taxinomie. Vol. 2: Analyse des Correspondances*. Paris: Dunod.
- Benzécri, J.-P. (1991). Measures, Models, and Graphical Displays in the Analysis of Cross-Classified Data: A Comment. *Journal of the American Statistical Association*, **86**(416), 1112-1115.
- Benzécri, J.-P. (1992). *Correspondence Analysis Handbook*. New York: Marcel Dekker, Inc.
- BMDP Statistical Software Inc. (1992). *BMDP Statistical Software Manual Release 7*, vols. 1 and 2. Los Angeles.
- Bond, J. & Michailidis, G. (1996). Homogeneity Analysis in Lisp-Stat. *Journal of Statistical Software*, 1(2). Διαθέσιμο στην ιστοσελίδα: <http://www.jstatsoft.org/v01/i02/PAPER/paper.html>.
- Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, **16**(3), 199-231.
- Burt, C. (1950). The Factorial Analysis of Qualitative Data. *British Journal of Psychology, Statistical Section*, **3**, 166-185.
- Chateau, F. & Lebart, L. (1996). Assessing Sample Variability in the Visualization Techniques Related to Principal Components Analysis: Bootstrap and Alternative Simulation Methods. *Proceedings: Computational Statistics COMPSTAT*, 205-210.
- Clausen, S.-E. (1998). *Applied Correspondence Analysis: An Introduction*. Thousand Oakes: Sage Publications.
- Cohen, J. & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. New Jersey: Lawrence Erlbaum Associates Inc.

- Davis, P. & Hersh, R. (1980). *The Mathematical Experience*. London: Penguin Books.
- De Leeuw, J. (1984). The GIFI-System of Non-Linear Multivariate Analysis. In E. Diday, M. Jambu, L. Lebart and R. Tomassone (Eds), *Data Analysis and Informatics III*, (pp.415-424). Amsterdam: North Holland.
- De Leeuw, J. (1993). Some Generalizations of Correspondence Analysis. In C. M. Cuadras and C. R. Rao (Eds), *Multivariate Analysis: Future Directions 2*, (pp. 359-375). Amsterdam: North-Holland.
- De Leeuw, J. (2005α). *Multivariate Analysis With Optimal Scaling*. Department of Statistics, UCLA. Department of Statistics Papers. Paper 2005103002. Διαθέσιμο στην ιστοσελίδα: <http://repositories.cdlib.org/uclastat/papers/2005103002>
- De Leeuw, J. (2005β). *Models and Techniques*. Department of Statistics, UCLA. Department of Statistics Papers. Paper 2005102703. Διαθέσιμο στην ιστοσελίδα: <http://repositories.cdlib.org/uclastat/papers/2005102703>
- De Leeuw, J., Young, F. & Takane, Y. (1976). Additive Structure in Qualitative Data: An Alternating Least Squares Methods with Optimal Scaling Features. *Psychometrika*, **41**(4), 471-503.
- Δερμάνης, Α. (1986). *Συνορθώσεις Παρατηρήσεων και Θεωρία Εκτίμησης, Τόμος 1*. Θεσσαλονίκη: Εκδόσεις Ζήτη.
- Deville J.-C. & Malinvaud, E. (1983). Data Analysis and Official Socio-Economic Statistics. *Journal of the Royal Statistical Society, Series A (General)*, **146**(4), 335-361.
- DiCiccio, T. & Efron, B. (1996). Bootstrap Confidence Intervals. *Statistical Science*, **11**(3), 189-228.
- Dillon, W. & Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications*. New York: John Willey & Sons, Inc.
- Efron, B. & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Fisher, R. A. (1940). The Precision of Discriminant Functions. *Annals of Eugenics*, **10**, 422-429.
- Gauch, H. G. Jr. (1995). *Multivariate Analysis in Community Ecology*. Cambridge: Cambridge University Press.
- Gifi, A. (1996). *Non-Linear Multivariate Analysis*. Chichester: John Willey & Sons Ltd.
- Gras, R. (1995). Ανάλυση ενός Ερωτηματολογίου με τη Συνεπαγωγική Μέθοδο. Στο Α. Γαγάτσης (Ed.), *Διδακτική και Ιστορία των Μαθηματικών*, (σ.σ. 97-109). Θεσσαλονίκη: ERASMUS ICP-94-G-2011/11.
- Greenacre, M. & Blasius, J. (Eds) (1994). *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*. London: Academic Press.
- Greenacre, M. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.
- Greenacre, M. (1993). *Correspondence Analysis in Practice*. London: Academic Press.
- Guttman, L. (1941). The Quantification of a Class of Attributes: A Theory and Method of Scale Construction. In P. Horst *et al.* (Eds), *The Prediction of Personal Adjustment*, (pp. 321-348). New York: Social Science Research Council.
- Hair, J., Anderson, R., Tatham, R. & Black, W. (1995). *Multivariate Data Analysis With Readings*. New Jersey: Prentice-Hall International, Inc.

- Heiser, W. & Meulman, J. (1994). Homogeneity Analysis: Exploring the Distribution of Variables and their Non-Linear Relationships. In M. Greenacre and J. Blasius (Eds), *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*, (pp. 179-209). London: Academic Press.
- Hillier, F. & Lieberman, G. (1995). *Introduction to Operations Research*. Singapore: McGraw-Hill, Inc.
- Hirschfeld, H. O. (1935). A Connection Between Correlation and Contingency. *Cambridge Philosophical Society Proceedings*, **31**, 520-524.
- Hoaglin, D. (2003). John W. Tukey and Data Analysis. *Statistical Science*, **18**(3), 311-318.
- Israëls, A. (1987). *Eigenvalue techniques for Qualitative Data*. Leiden: DSWO Press.
- Johnson, D. (1998). *Applied Multivariate Methods for Data Analysis*. Pacific Grove: Brook/Cole Publishing Company.
- Johnson, R. & Wichern, D. (1992). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice – Hall, Inc.
- Καραπιστόλης, Δ. (1999). *Ανάλυση Δεδομένων και Έρευνα Αγοράς*. Θεσσαλονίκη: Εκδόσεις ANIKOYΛA.
- Kruskal, J. & Shepard, R. (1974). A Nonmetric Variety of Linear Factor Analysis. *Psychometrika*, **39**(2), 123-157.
- Kruskal, J. & Wish, M. (1978). *Multidimensional Scaling*. Newbury Park: Sage Publications.
- Le Roux. B. & Rouanet, H. (2004). *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*. Dordrecht: Kluwer Academic Publishers.
- Lebart, L. (2005). Validation Techniques in Simple and Multiple Correspondence Analysis. *Τετράδια Ανάλυσης Δεδομένων-Data Analysis Bulletin*, **6/05**, 10-25.
- Lebart, L., Morineau, A. & Warwick, K. M. (1984). *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. New York: John Willey, Inc.
- 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).
- Μεϊμάρης, Μ. (2002). Εξισώσεις Μεταφοράς και Μεταφορά Εξισώσεων. *Τετράδια Ανάλυσης Δεδομένων-Data Analysis Bulletin*, **1**, 12-19.
- Μενεξές, Γ. (2007). *Πειραματικοί Σχεδιασμοί στην Ανάλυση Δεδομένων*. Διδακτορική Διατριβή που υποβλήθηκε στο Τμήμα Εφαρμοσμένης Πληροφορικής του Πανεπιστημίου Μακεδονίας.
- Μενεξές, Γ. (2008). Μια Δομημένη Προσέγγιση στην Πολυμεταβλητή Στατιστική Ανάλυση Βιολογικών, Περιβαλλοντικών, Κοινωνικών και Οικονομικών Δεδομένων. Στο Αραμπατζής, Γ. και Πολύζος, Σ. (Eds), *Φυσικοί Πόροι, Περιβάλλον & Ανάπτυξη*, (σ.σ. 519-534), Θεσσαλονίκη: Εκδόσεις Τζιόλα.
- Meulman J. & Heiser, W. (2004). *SPSS Categories 13.0*. Chicago: SPSS Inc.
- Meulman, J., (1999). *Optimal Scaling Methods for Multivariate Categorical Data Analysis*. SPSS White Paper. Διαθέσιμο στην ιστοσελίδα: http://www.spss.com/cool/papers/optimal_scaling.html
- Mertens, D. (1998). *Research Methods in Education and Psychology: Integrating Diversity with Quantitative & Qualitative Approaches*. Thousand Oaks, Sage Publications, Inc.

- Michailidis, G. & De Leeuw, J. (1998). The Gifi System of Descriptive Multivariate Analysis. *Statistical Science*, **13**(4), 307-336.
- Michailidis, G. & De Leeuw, J. (2000). Multilevel Homogeneity Analysis With Differential Weighting. *Computational Statistics & Data Analysis*, **32**, 411-442.
- Michailidis, G. & De Leeuw, J. (2005). Homogeneity Analysis Using Absolute Deviations. *Computational Statistics & Data Analysis*, **48**, 587-603.
- Michailidis, G. (1996). *Multilevel Homogeneity Analysis*. Διδακτορική διατριβή που υποβλήθηκε στο Τμήμα Μαθηματικών του Πανεπιστημίου της Καλιφόρνια στο Λος Άντζελες.
- Moran, M. A. & Gornbein, J. (1988). *CA-Correspondence Analysis*. Technical Report #87, BMDP Statistical Software, Inc., Los Angeles.
- Μπερχράκης, Θ. (1999). *Πολυδιάστατη Ανάλυση Δεδομένων: Μέθοδοι και Εφαρμογές*, Αθήνα: Εκδόσεις ΝΕΑ ΣΥΝΟΡΑ-Α. Α. ΛΙΒΑΝΗΣ.
- Nishisato, S. (1978). Optimal Scaling of Paired Comparison and Rank Order Data: An Alternative to Guttman's Formulation. *Psychometrika*, **43**(2), 263-271.
- Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and its Applications*. Toronto: University of Toronto Press.
- Nishisato, S. (1993). On Quantifying Different Types of Categorical Data. *Psychometrika*, **58**(4), 617-629.
- Nishisato, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Nishisato, S. (1996). Gleaning in the Field of Dual Scaling. *Psychometrika*, **61**(4), 559-599.
- Norusis, M. (1992). *SPSS Professional Statistics 6.1*. Chicago: SPSS Inc.
- Παπαδημητρίου, Γ. (1994). *Μέθοδοι Ανάλυσης Δεδομένων: Πανεπιστημιακές Παραδόσεις*. Θεσσαλονίκη: Έκδοση Πανεπιστημίου Μακεδονίας Οικονομικών και Κοινωνικών Επιστημών
- Παπαδημητρίου, Γ. (2002). Η Ανάλυση Δεδομένων στην Ελλάδα. *Τετράδια Ανάλυσης Δεδομένων-Data Analysis Bulletin*, **1/02**, 5-11.
- Παπαδημητρίου, Γ. (2004). *Πολυμεταβλητή Στατιστική Ανάλυση: Πανεπιστημιακές Παραδόσεις*. Θεσσαλονίκη: Έκδοση Πανεπιστημίου Μακεδονίας Οικονομικών και Κοινωνικών Επιστημών.
- Παπαδημητρίου, Γ. (2007). *Η Ανάλυση Δεδομένων. Παραγοντική Ανάλυση των Αντιστοιχιών, Ιεραρχική Ταξινόμηση και άλλες Μέθοδοι*. Αθήνα: Εκδόσεις τυπωθήτω, Γ. Δαρδανός.
- Pfeiffer, P. (1978). *Concepts of Probability*. New York: Dover Publications, Inc.
- SAS Institute, (1999). *SAS/STAT User's Guide Version 8*. Cary, NC: SAS Institute, Inc.
- Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Wiley & Sons, Inc.
- Stevens, J. (2002). *Applied Multivariate Statistics for the Social Sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Tacq, J. (1997). *Multivariate Analysis Techniques in Social Science Research*. London: Sage Publications.
- Takane, Y., Young, F. & De Leeuw, J. (1977). Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Methods With Optimal Scaling Features. *Psychometrika*, **42**(1), 7-67.

- Tenenhaus, M. & Young, F. (1985). An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and Other Methods for Quantifying Categorical Multivariate Data. *Psychometrika*, **50**(1), 91-119.
- Tukey, J. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, **33**(1), 1-67.
- Tukey, J. (1977). *Exploratory Data Analysis*. Menlo Park: Addison-Wesley Publishing Company.
- Tukey, J. (1980). We Need Both Exploratory and Confirmatory. *The American Statistician*, **34**(1), 23-25.
- Van de Geer, J. P. (1993 α). *Multivariate Analysis of Categorical Data: Theory*. Thousand Oakes: Sage Publications, Inc.
- Van de Geer, J., (1993 β). *Multivariate Analysis of Categorical Data: Applications*. Thousand Oakes: Sage Publications, Inc.
- Van der Burg, E. & De Leeuw, J. & Dijksterhuis, G. (1994). OVERALS: Non Linear Canonical Correlation with k Sets of Variables. *Computational Statistics & Data Analysis*, **18**, 141-163.
- Van der Burg, E. & De Leeuw, J. (1983). Non-linear Canonical Correlation. *British Journal of Mathematical and Statistical Psychology*, **36**, 54-80.
- Van der Burg, E. & De Leeuw, J. (1988). Homogeneity Analysis with k sets of Variables: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, **53**(2), 177-197.
- Van Meter, K., Schiltz, M.-A., Cibois, F. & Mounier, L. (1994). Correspondence Analysis: A History and French Sociological Perspective, In J. Blasius and M. Greenacre (Eds), *Visualization of Categorical Data*, (pp. 128-137). San Diego: Academic Press.
- Van Rijckeversel, J. & De Leeuw, J. (Eds) (1988). *Component and Correspondence Analysis. Dimension Reduction by Functional Approximation*, (pp. 103-114). Chichester: John Wiley & Sons Ltd.
- Van Rijckeversel, J. (1987). *The Application of Fuzzy Coding and Horseshoes in Multiple Correspondence Analysis*. Leiden: DSWO Press.
- Weller, S. & Romney A. K. (1990). *Metric Scaling: Correspondence Analysis*. Newbury Park: Sage Publications.
- Young, F. (1981). Quantitative Analysis of Qualitative Data. *Psychometrika*, **46**(4), 357-388.
- Young, F., De Leeuw, J. & Takane, Y. (1976). Regression With Qualitative and Quantitative Variables: An Alternating Least Squares Methods With Optimal Scaling Features. *Psychometrika*, **41**(4), 505-529.
- Young, F., Takane, Y. & De Leeuw, J. (1978). The Principal Components of Mixed Measurement Level Multivariate Data: An Alternating Least Squares Methods With Optimal Scaling Features. *Psychometrika*, **43**(2), 279-281.