



Πολυμεταβλητή Στατιστική Ανάλυση

Δρ. Αθανάσιος Δαγούμας,
Επ. Καθηγητής Οικονομικής της
Ενέργειας & των Φυσικών
Πόρων, Πανεπιστήμιο Πειραιώς

Πολυμεταβλητή ανάλυση

Στην πράξη τα **δεδομένα** ενός ερευνητή είναι από τη φύση τους **πολυμεταβλητά**

Σπάνια μια μεταβλητή μελετάται ανεξάρτητα και απομονωμένα από τις υπόλοιπες

Η **στατιστική συμπερασματολογία με τη χρήση πολλών μεταβλητών** ονομάζεται **πολυμεταβλητή στατιστική ανάλυση**



Πολυμεταβλητή ανάλυση

Διαδεδομένες πολυμεταβλητές μέθοδοι:

- Η **πολλαπλή γραμμική παλινδρόμηση** (και επομένως και το γενικό γραμμικό μοντέλο - linear regression model) στοχεύει στη μελέτη πολλών μεταβλητών συγχρόνως.
- **Ανάλυση διακύμανσης** (Multivariate Analysis Of Variance - ANOVA)



Πολυμεταβλητή ανάλυση

Είναι χρήσιμη διότι:

- Έχουμε **περισσότερη πληροφορία** (περισσότερες μεταβλητές ερμηνεύουν καλύτερα το φαινόμενο)
 - Περιορίζεται η αβεβαιότητα
 - Εξάγει συμπεράσματα με μεγαλύτερη βαρύτητα
- Μελετάμε **συσχετισμούς** (μεταξύ μεταβλητών και μεταξύ οντοτήτων/υποκειμένων).
 - Η ανακάλυψη συσχετίσεων και συνδιακυμάνσεων μεταξύ των μεταβλητών απλοποιεί την ανάλυση και την εξαγωγή συμπερασμάτων



Πολυμεταβλητή ανάλυση

Χρησιμοποιείται για:

- Την **εύρεση και ερμηνεία συσχετίσεων** μεταξύ των μεταβλητών.
- Τη **δημιουργία ομάδων** είτε από παρατηρήσεις είτε από μεταβλητές σύμφωνα με κάποια χαρακτηριστικά
 - Ομαδοποίηση (clustering)
- Τη **μείωση των διαστάσεων** του προβλήματος
 - συμπύκνωση της πληροφορίας σε λιγότερες μεταβλητές
- Την **πρόβλεψη νέων τιμών**
 - εκτίμηση τιμών σε χαμένες παρατηρήσεις (missing observations).
 - κατασκευάζουμε κανόνες ώστε να μπορούμε να κατατάσσουμε νέες παρατηρήσεις σε ομάδες



Πολυμεταβλητή ανάλυση

Χρησιμοποιείται για:

- **Μοντελοποίηση σε πολλές διαστάσεις**
 - Στη γραμμική παλινδρόμηση υπάρχει μια εξαρτημένη και πολλές ανεξάρτητες μεταβλητές
 - Υπάρχουν μοντέλα πολυμεταβλητής παλινδρόμησης με περισσότερες εξαρτημένες μεταβλητές
- **Ποσοτικοποίηση μη παρατηρήσιμων ποσοτήτων**
 - Αφηρημένες και μη μετρήσιμες έννοιες (π.χ. ευφύια, αλληλεγγύη)
 - Ο συνδυασμός άλλων μετρήσιμων μεταβλητών οδηγούν στην ποσοτικοποίηση τέτοιων εννοιών (εμπεριέχεται όμως κάποιος βαθμός αυθαιρεσίας και επισφάλειας)



Πολυμεταβλητή ανάλυση

Κύριες πολυμεταβλητές τεχνικές:

- **Ανάλυση σε κύριες συνιστώσες (Principal Components Analysis)**
 - ερμηνεύει συσχετίσεις ανάμεσα στις αρχικές μεταβλητές,
 - μειώνει τις διαστάσεις του προβλήματος (συνιστώσες)
 - στην πράξη είναι απλά ένας μαθηματικός μετασχηματισμός των δεδομένων, όπου
 - βρίσκει γραμμικούς συνδυασμούς των αρχικών δεδομένων έτσι ώστε η πληροφορία να μη χάνεται
 - άλλα οι νέες μεταβλητές, οι συνιστώσες, είναι ασυσχέτιστες μεταξύ τους



Πολυμεταβλητή ανάλυση

Κύριες πολυμεταβλητές τεχνικές:

- **Παραγοντική ανάλυση (Factor Analysis)**

- εύρεση και ερμηνεία παραγόντων που δεν είναι μετρήσιμοι, αλλά υπάρχουν και προκαλούν τη συσχέτιση μεταξύ των παρατηρούμενων μεταβλητών.
- σε αντίθεση με την ανάλυση σε κύριες συνιστώσες, έχει ένα ισχυρό θεωρητικό υπόβαθρο και επομένως επιτρέπει τη στατιστική εξέταση διάφορων υποθέσεων σχετικά με το υπό μελέτη φαινόμενο.



Πολυμεταβλητή ανάλυση

Κύριες πολυμεταβλητές τεχνικές:

- **Ανάλυση σε ομάδες (Cluster Analysis)**
 - δημιουργία ομάδων (clusters) από παρατηρήσεις για τις οποίες τα δεδομένα δείχνουν πως έχουν παρόμοια χαρακτηριστικά.
 - είναι γνωστή ως Ταξινόμηση (Taxonomy), Κατηγοριοποίηση (Classification) αλλά και ως Segmentation
 - Η ταξινόμηση γίνεται με διάφορες μεθόδους:
 - εμπειρικές, δηλαδή χωρίς σημαντικό θεωρητικό υπόβαθρο, και βασίζονται στην έννοια της απόστασης
 - βασισμένες σε μοντέλα (model-based clustering), οι οποίες αφενός έχουν ένα σημαντικό θεωρητικό υπόβαθρο, αφετέρου προσφέρουν μια σειρά από μεθοδολογικά εργαλεία, για να μπορεί κανείς να αξιολογήσει τα αποτελέσματα.



Πολυμεταβλητή ανάλυση

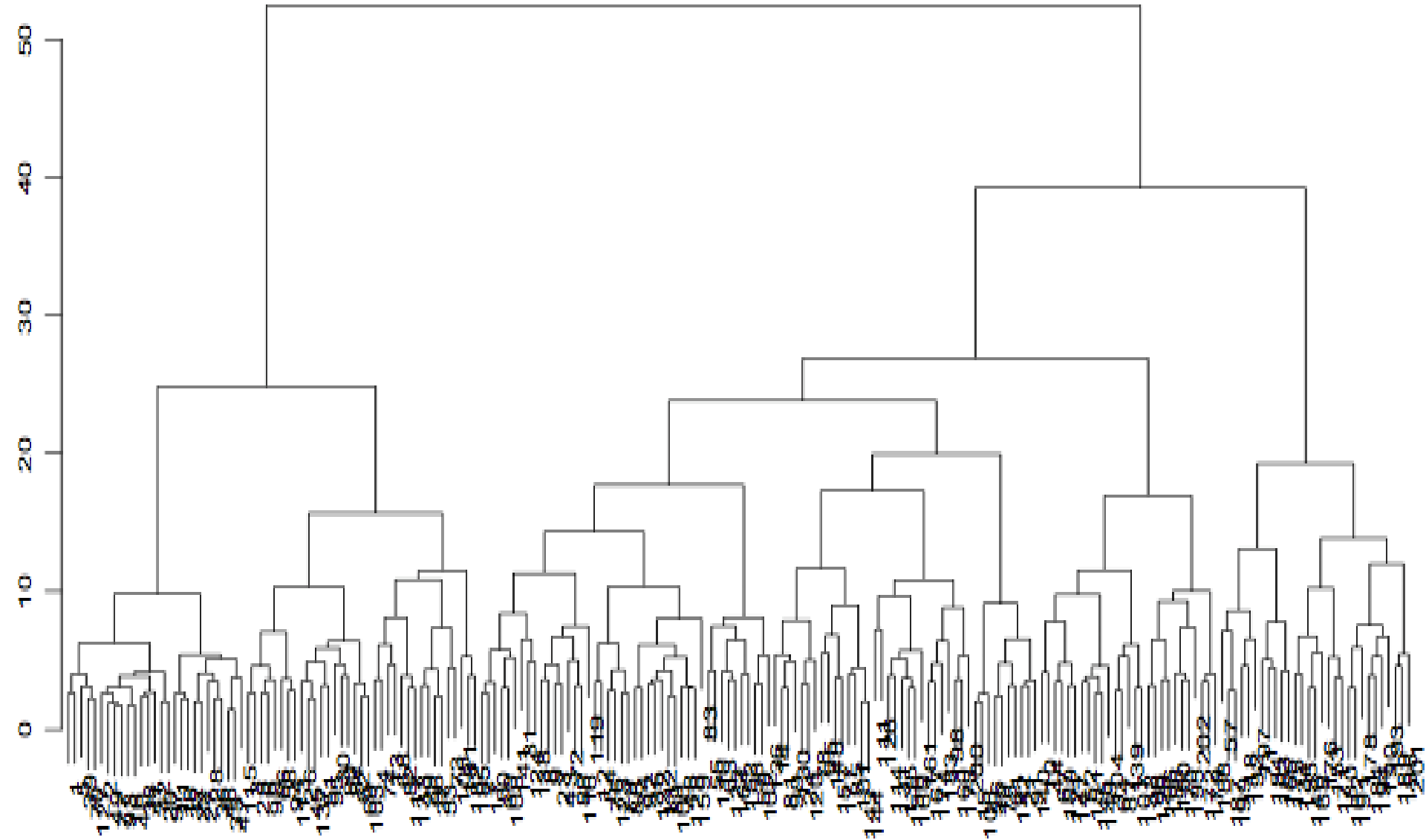
Κύριες πολυμεταβλητές τεχνικές:

- **Ανάλυση σε ομάδες (Cluster Analysis)**

- Με Ιεραρχικούς Αλγορίθμους διαμελίζουμε έναν πληθυσμό παρατηρήσεων σε έναν αριθμό ομάδων, που είναι ομοιογενείς
- Στόχος της μεθόδου να δημιουργήσει όσο δυνατόν πιο εσωτερικά συμπαγείς ομάδες και ταυτόχρονα ανομοιογενείς μεταξύ τους ομάδες
- Το δένδρογραμμα δημιουργείται με τις διαδοχικές συνενώσεις, αρχικά 2 πιο όμοιων παρατηρήσεων και στη συνέχεια των ομάδων, μέχρι να φθάσουμε στον αρχικό κόμβο, δηλαδή τη μέση παρατήρηση του πληθυσμού που περιλαμβάνει το σύνολο των αρχικών παρατηρήσεων



Δενδρόγραμμα



Πολυμεταβλητή ανάλυση

Κύριες πολυμεταβλητές τεχνικές:

- **Διακριτική ανάλυση (Discriminant Analysis)**
 - δημιουργήσει κανόνες από ήδη υπάρχοντα δεδομένα, ώστε να είναι σε θέση κανείς να κατατάξει μελλοντικές παρατηρήσεις σε έναν από τους υπό εξέταση πληθυσμούς.
 - είναι γνωστή ως Αναγνώριση Προτύπων (Pattern Recognition) και αποσκοπεί να δημιουργήσει κανόνες που θα μπορούν να ξεχωρίζουν συγκεκριμένα πρότυπα (π.χ. κείμενα, εικόνες κ.λπ.) από ένα μεγάλο όγκο πληροφορίας.



Πολυμεταβλητή ανάλυση

Κύριες πολυμεταβλητές τεχνικές:

- **Ανάλυση αντιστοιχιών (Correspondence Analysis)**
 - μέθοδος συγγενής της ανάλυσης σε κύριες συνιστώσες
 - η συλλογή των στοιχείων γίνεται με τη χρήση ερωτηματολογίου και άρα τα δεδομένα είναι από τη φύση τους κατηγορικά.
 - Στόχος είναι να δημιουργηθούν άξονες πάνω στους οποίους προβάλλει κανείς τις παρατηρήσεις αλλά και τις μεταβλητές και επομένως μπορεί να δει και να ερμηνεύσει πώς σχετίζονται μεταξύ τους οι μεταβλητές.



Πολυμεταβλητή ανάλυση

Κύριες πολυμεταβλητές τεχνικές:

- **Ανάλυση κανονικών συσχετίσεων (Canonical Correlation Analysis)**
 - παρόμοια με την ανάλυση σε κύριες συνιστώσες
 - η διαφορά είναι ότι οι συνιστώσες που προκύπτουν έχουν μεταξύ τους κάποια ελεγχόμενη συσχέτιση.
 - με αυτό τον τρόπο επιτυγχάνουμε μια πιο ρεαλιστική περιγραφή των δεδομένων.



Πολυμεταβλητή ανάλυση

Κύριες πολυμεταβλητές τεχνικές:

- **Πολυδιάστατη κλιμακοποίηση (Multidimensional Scaling)**
 - μαθηματική μέθοδος που σκοπό έχει να προβάλλει τις διαστάσεις του προβλήματος στο χώρο των δύο (συνήθως) ή περισσότερων διαστάσεων.
 - είναι πολύ πιο εύκολο να μελετήσουμε ένα διάγραμμα λίγων διαστάσεων σε σχέση με δεδομένα πολλών διαστάσεων χωρίς ουσιαστικά κανένα εργαλείο απεικόνισής τους
 - καταφέρνει να δημιουργήσει δείκτες βασισμένους σε όλα τα δεδομένα οι οποίοι είναι πιο εύκολα κατανοητοί.



Πολυμεταβλητή ανάλυση

Κύριες πολυμεταβλητές τεχνικές:

- **Πολυμεταβλητό Γραμμικό Μοντέλο (Multivariate Linear Model)**
 - περιλαμβάνει την παλινδρόμηση και την ανάλυση διακύμανσης ως ειδικές περιπτώσεις
 - υπάρχει μια εξαρτημένη μεταβλητή και πολλές συνήθως ανεξάρτητες, οι οποίες στην περίπτωση της ανάλυσης διακύμανσης είναι κατηγορικές.
 - το μοντέλο αυτό αν γενικευτεί στις πολλές διαστάσεις, υπάρχουν πολλές εξαρτημένες μεταβλητές, οι οποίες μπορεί να έχουν και μεταξύ τους κάποια συσχέτιση.
 - έτσι προκύπτει η μέθοδος της πολυμεταβλητής παλινδρόμησης (Multivariate Regression) και η μέθοδος της πολυμεταβλητής ανάλυσης διακύμανσης (MANOVA)



Πολυμεταβλητή ανάλυση

Κύριες πολυμεταβλητές τεχνικές:

- **Μέθοδοι για δεδομένα διεύθυνσης (Directional data)**
 - πολλές παρατηρήσεις σχετίζονται με την έννοια της διεύθυνσης (π.χ. ανεμολογικά δεδομένα)
 - όταν η διεύθυνση είναι μια μεταβλητή, υπάρχουν και διάφορες άλλες μεταβλητές που περιγράφουν το φαινόμενο (π.χ. ένταση ανέμου, γεωγραφικές διαστάσεις, ύψος).
 - υπάρχουν ειδικές μέθοδοι για τέτοιου είδους πολυμεταβλητά δεδομένα



Συμπεράσματα

Τα περισσότερα **δεδομένα** που χρησιμοποιεί ένας ερευνητής είναι **πολυμεταβλητά**

Για την **καλύτερη αξιοποίηση των δεδομένων**, είναι πολύ χρήσιμη:

- η εύρεση και ερμηνεία των συσχετίσεων μεταξύ των μεταβλητών
- η ομαδοποίηση των παρατηρήσεων
- η μείωση των διαστάσεων του προβλήματος
- η πρόβλεψη νέων τιμών (τιμές για αυτές που λείπουν ή για τις ομάδες που πιθανόν να ανήκουν)
- η μοντελοποίηση σε πολλές διαστάσεις (πολλές εξαρτημένες μεταβλητές)
- η ποσοτικοποίηση μη παρατηρήσιμων ποσοτήτων



Συμπεράσματα

Κύριες μέθοδοι πολυμεταβλητής ανάλυσης:

- Ανάλυση σε κύριες συνιστώσες (Principal Components Analysis)
- Παραγοντική ανάλυση (Factor Analysis)
- Ανάλυση σε ομάδες (Cluster Analysis)
- Διακριτική ανάλυση (Discriminant Analysis)
- Ανάλυση αντιστοιχιών (Correspondence Analysis)
- Ανάλυση κανονικών συσχετίσεων (Canonical Correlation Analysis)
- Πολυδιάστατη κλιμακοποίηση (Multidimensional Scaling)
- Πολυμεταβλητό Γραμμικό Μοντέλο (Multivariate Linear Model)
- Μέθοδοι για δεδομένα διεύθυνσης (Directional data)



Συμπεράσματα

Οι μέθοδοι πολυμεταβλητής ανάλυσης
κατηγοριοποιούνται σε **2 γενικές κατηγορίες**:

Μέθοδοι Ποσοτικοποίησης & Μέθοδοι Ταξινόμησης

Overview of Multivariate Analysis

Quantitation Group

Multiple regression
Principal component regression (PCR)
Partial least squares (PLS)

Classification Group

Principal component analysis (PCA)
Cluster analysis
Discriminant analysis
Factor analysis
... etc.

Fig. 1 Overview of Multivariate Analysis