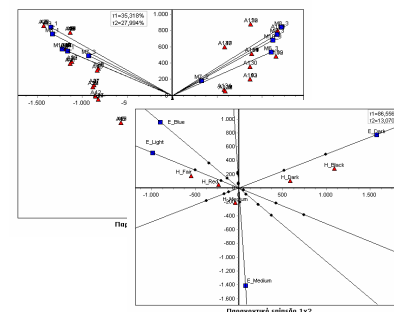
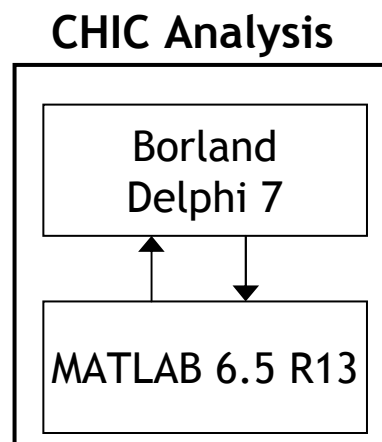


Η Παραγοντική Ανάλυση των Αντιστοιχιών μέσω του λογισμικού CHIC Analysis

Άγγελος Μάρκος, Γεώργιος Μενεξές, Γιάννης Παπαδημητρίου
Τμήμα Εφαρμοσμένης Πληροφορικής, Πανεπιστήμιο Μακεδονίας

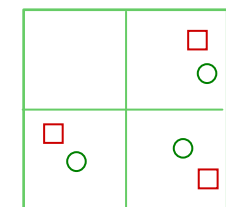
- Το C.H.I.C. (*Correspondence & Hierarchical Cluster*) Analysis αποτελεί ένα ολοκληρωμένο λογισμικό εφαρμογής της διμεταβλητής και πολυμεταβλητής Παραγοντικής Ανάλυσης των Αντιστοιχιών (Π.Α.Α.)
- Συνδυάζει τις δυνατότητες γραφικών μιας γλώσσας οπτικού προγραμματισμού (Delphi), με τις υπολογιστικές δυνατότητες μιας μαθηματικής γλώσσας προγραμματισμού (MATLAB).
- Ενσωματώνει τα αριθμητικά και γραφικά αποτελέσματα αρκετών εμπορικών και μη λογισμικών (SPSS, SAS, Minitab, SPAD, Statistica κ.α.) και παρέχει νέες δυνατότητες και δείκτες.



```

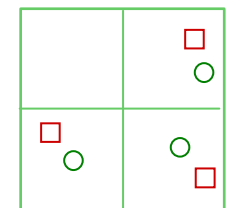
124 % Source of chi-square cell contributions
125 %
126 % = (O - expected) * expected / (expected * total)
127
128 % Direction of row and column
129 row = sum(P.^2);
130 col = sum(P.^2);
131
132 % Chi-square statistic
133 chiqrstat = sum((O - expected).^2) / expected;
134
135 % P-value
136 p1 = 1 - gamcdf(chiqrstat, df/2, 2);
137
138 % Chi-square contributions
139 chiqrcon = (O - expected) ./ expected;
140
141 % Solve using compressed form of single value decomposition (this SVD)
142 if n>=
143     [U,S,V] = svd(P,0);
144 else
145     [U,S,V] = svd(P,0);
146 end
147
148 % Keep positive ones
149
150

```



Δυνατότητες - Χαρακτηριστικά

- Α.Βοήθεια στην ερμηνεία των αποτελεσμάτων της μεθόδου
- Β.Δυνατότητα ανάλυσης υποσυνόλων του γενικευμένου πίνακα συμπτώσεων (πίνακα Burt)
- Γ.Δυνατότητα ανάλυσης μεγάλων συνόλων δεδομένων με τη χρήση αποτελεσματικού αλγορίθμου.
- Δ.Δυνατότητες αλληλεπίδρασης με τις διαγραμματικές εκροές της μεθόδου



Α. Βοήθεια στην ερμηνεία των αποτελεσμάτων Διμεταβλητή Περίπτωση

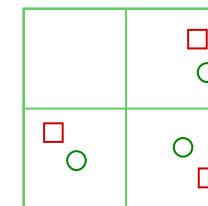
1. Κριτήρια σημαντικότητας σημείων επί παραγοντικών αξόνων και επιπέδων

Ενεργά σημεία

- MASS - Βάρος ή Μάζα
- INR - Αδράνεις σημείων γραμμών και στηλών
- QLT - Ποιότητα απεικόνισης
- COR - Correlation²
- SQCOR- Τετραγωνική ρίζα του COR (Blasius & Greenacre, 1994)
- CTR - Contribution (Απόλυτη και Σχετική)
- PF (Παπαδημητρίου & Φλώρου, 1994)
- Best (SAS)

Συμπληρωματικά σημεία

- INR - Αδράνεις σημείων γραμμών και στηλών
- QLT - Ποιότητα απεικόνισης
- COR - Correlation²



Α. Βοήθεια στην ερμηνεία των αποτελεσμάτων Διμεταβλητή Περίπτωση

2. Κριτήρια προσδιορισμού του αριθμού των παραγόντων που χρήζουν ερμηνείας

2.1 Μελέτη των Ιδιοτιμών

- Διάγραμμα Ιδιοτιμών (Scree Plot) (SPSS)

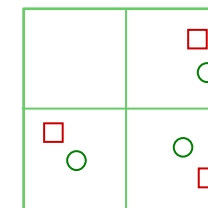
2.2 Στατιστική σημαντικότητα παραγοντικών αξόνων

- Κριτήριο Nishisato (1980)
- Κριτήριο Van de Geer (1993)

3. Στατιστική σημαντικότητα της ολικής αδράνειας

4. Επιπλέον Δυνατότητες - Δείκτες

- Αναμενόμενη Συχνότητα
- Υπόλοιπα / Τυποποιημένα Υπόλοιπα / Διορθωμένα Τυποποιημένα Υπόλοιπα
- Συνεισφορά ιδιοτήτων στο χ^2
- Συνεισφορά αξόνων στο χ^2
- Πίνακας Ανασύστασης των Αρχικών Δεδομένων



Α. Βοήθεια στην ερμηνεία των αποτελεσμάτων Πολυμεταβλητή Περίπτωση

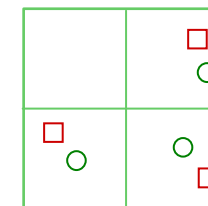
1. Κριτήρια σημαντικότητας σημείων επί παραγοντικών αξόνων και επιπέδων

Ενεργά σημεία

- MASS - Βάρος ή Μάζα
- INR - Αδράνεις σημείων γραμμών και στηλών
- QLT - Ποιότητα απεικόνισης
- COR - Correlation²
- SQCOR - Τετραγωνική ρίζα του COR (Blasius & Greenacre, 1994)
- CTR - Contribution (Απόλυτη και Σχετική)
- Άθροισμα CTR ανά μεταβλητή (SPAD)
- PF (Παπαδημητρίου & Φλώρου, 1994)
- Best (SAS)
- Μέτρα Διακριτότητας (Discrimination Measures) (SPSS) / Σχετικά Μέτρα Διακριτότητας

Συμπληρωματικά σημεία

- INR - Αδράνεις σημείων γραμμών και στηλών
- QLT - Ποιότητα απεικόνισης
- COR - Correlation²
- T-Τιμές (SPAD)



Α. Βοήθεια στην ερμηνεία των αποτελεσμάτων Πολυμεταβλητή Περίπτωση

2. Κριτήρια προσδιορισμού του αριθμού των παραγόντων που χρήζουν ερμηνείας

2.1 Μελέτη των Ιδιοτιμών

- Διάγραμμα Ιδιοτιμών (Scree Plot)
- Παράγοντες με ιδιοτιμή $> 1/p$ και $1/p^2$ (p : αριθμός μεταβλητών)
- Παράγοντες με Cronbach's $\alpha > 0$

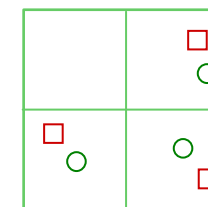
2.2 Στατιστική σημαντικότητα παραγοντικών αξόνων

- Κριτήριο Nishisato (1980)
- Διορθωμένο Κριτήριο του Kaiser (Μενεξές, Μάρκος & Παπαδημητρίου, 2006)

3. Στατιστική σημαντικότητα της «ενδιαφέρουσας αδράνειας»

4. Διόρθωση των Ιδιοτιμών

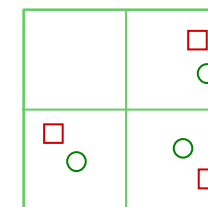
- Διόρθωση κατά Greenacre (1993)
- Διόρθωση σύμφωνα με «ενδιαφέρουσα αδράνεια» (Μενεξές, Μάρκος & Παπαδημητρίου, 2006)



A. Βοήθεια στην ερμηνεία των αποτελεσμάτων

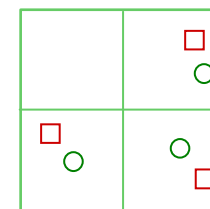
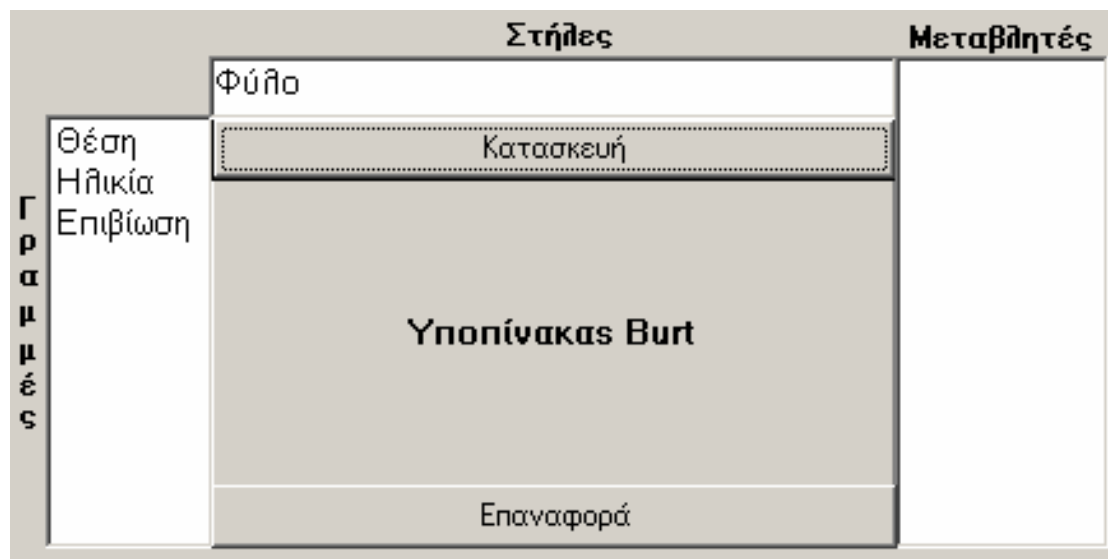
5. Μέθοδοι Κανονικοποίησης των Συντεταγμένων (SPSS)

- α) Κύρια Κανονικοποίηση κατά γραμμές (RPN - Row Principal Normalization)
- β) Κύρια Κανονικοποίηση κατά στήλες (CPN - Row Principal Normalization)
- γ) Συμμετρική Κανονικοποίηση (SN - Symmetrical Normalization)
- δ) Κύρια Κανονικοποίηση (PN - Principal Normalization)



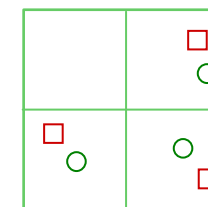
Β. Δυνατότητα ανάλυσης υποσυνόλων του πίνακα Burt

- Αλγόριθμος εντοπισμού του «καλύτερου» υποπίνακα σύμφωνα με το κριτήριο της «ενδιαφέρουσας αδράνειας» (Μάρκος, Μενεξές & Παπαδημητρίου, 2005)
- Εργαλείο δυναμικής κατασκευής υποπινάκων του γενικευμένου πίνακα συμπτώσεων (Burt)



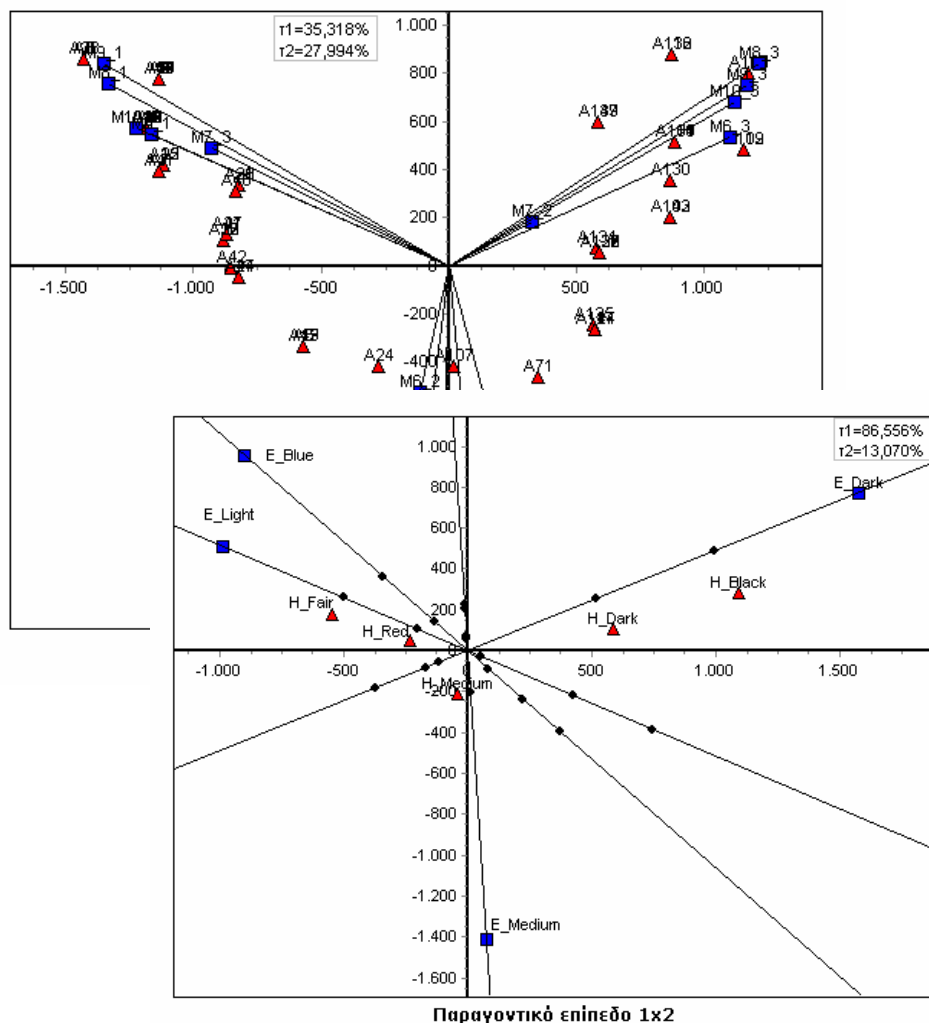
Γ. Δυνατότητα ανάλυσης μεγάλων συνόλων δεδομένων με τη χρήση αποτελεσματικού αλγορίθμου

- Το CHIC Analysis υλοποιεί μια αποτελεσματική εκδοχή του αλγορίθμου της Π.Α.Α. που καθιστά εφικτή την ανάλυση λογικών πινάκων με μεγάλο αριθμό γραμμών (αντικειμένων) (Μενεξές & Παπαδημητρίου, 2004).
- Το υπολογιστικό κέρδος είναι σημαντικό: α) 45% κατά μέσο όρο ταχύτερος αλγόριθμος, β) δυνατότητα ανάλυσης πινάκων με 500.000 γραμμές (αντικείμενα) και 120 στήλες (κλάσεις μεταβλητών).

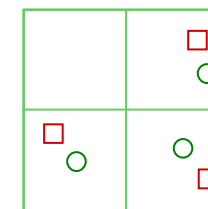


Δ. Δυνατότητες αλληλεπίδρασης γραφικών

Παραγοντικοί άξονες (1D & 2D)



- Επεξεργασία γραφημάτων (κλίμακα, τίτλος, σύμβολα, μετακίνηση ετικετών)
- Πληροφορίες για κάθε σημείο (querying)
- Biplot Άξονες (διμεταβλητή περίπτωση)
- Διανύσματα θέσης (πολυμεταβλητή περίπτωση)



- **Σύνολο δεδομένων**

Titanic Data Set (2201x4) (Dawson, 1995)

- **Περιγραφή**

Αντικείμενα: 2201 επιβάτες του «Τιτανικού»

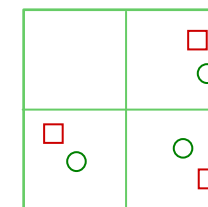
Μεταβλητές:

M1: Θέση [1.α' θέση 2.β' θέση 3.γ' θέση 4.πλήρωμα]

M2: Ηλικία [1.παιδί 2.ενήλικος]

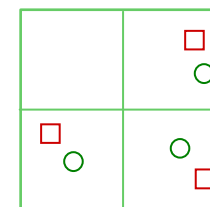
M3: Φύλο [1.γυναίκα 2.άντρας]

M4: Επιβίωση [1.δεν επιβίωσε 2.επιβίωσε]



Συμπεράσματα

- Το λογισμικό **CHIC Analysis** διαθέτει:
 - α) ποικιλία δεικτών, υπαρχόντων και νέων
 - β) αποτελεσματικό αλγόριθμο για την ανάλυση μεγάλων πινάκων δεδομένων
 - γ) δυνατότητες αλληλεπίδρασης με τα γραφικά αποτελέσματα



Πληροφορίες

Το CHIC Analysis διανέμεται ελεύθερα στο διαδίκτυο:

<http://datan.uom.gr/amarkos/analysis.html>

