ΠΑΝΕΠΙΣΤΗΜΙΟ ΣΤΕΡΕΑΣ ΕΛΛΑΔΟΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΪΑΤΡΙΚΗ

Διαταραχή ιδιοτιμών συμμετρικού πίνακα

Ειρήνη Καλαθά

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Επιβλέποντες
Μαρία Αδάμ
Παντελεήμων Μπάγκος
Επίκουροι Καθηγητές

Λαμία, Ιούλιος 2012

ΠΑΝΕΠΙΣΤΗΜΙΟ ΣΤΕΡΕΑΣ ΕΛΛΑΔΟΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΪΑΤΡΙΚΗ

Διαταραχή ιδιοτιμών συμμετρικού πίνακα

Ειρήνη Καλαθά

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
Επιβλέποντες
Μαρία Αδάμ
Παντελεήμων Μπάγκος
Επίκουροι Καθηγητές

Λαμία, Ιούλιος 2012

Διαταραχή ιδιοτιμών συμμετρικού πίνακα

Ειρήνη Καλαθά

Τριμελής Επιτροπή:

Μαρία Αδάμ, Επίκουρος Καθηγήτρια του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική (επιβλέπουσα)

Παντελεήμων Μπάγκος, Επίκουρος Καθηγητής του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική (επιβλέπων)

Βασίλειος Πλαγιανάκος, Επίκουρος Καθηγητής του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική

Ευχαριστίες

Η παρούσα πτυχιακή εργασία πραγματοποιήθηκε στο Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Στερεάς Ελλάδος από τον Οκτώβριο του 2010 έως το Ιούλιο του 2012.

Έχοντας ολοκληρώσει την πτυχιακή εργασία μου, θα ήθελα να ευχαριστήσω θερμά:

Τους επιβλέποντες καθηγητές μου, κα. Μαρία Αδάμ και κ. Παντελή Μπάγκο, Επίκουρους Καθηγητές του Τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Στερεάς Ελλάδος, για τη συνεχή παρακολούθηση, υποστήριξη και ενθάρρυνση κατά τη διάρκεια της εκπόνησης της παρούσας εργασίας. Όπως επίσης, και για την πολύτιμη βοήθεια και καθοδήγησή τους, για την επίλυση των διάφορων θεμάτων καθ' όλη την πορεία της ανάπτυξης και της συγγραφής αυτής της εργασίας.

Τον κ. Βασίλειο Πλαγιανάκο, μέλος της Τριμελούς Επιτροπής και Επίκουρος Καθηγητής στο Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Στερεάς Ελλάδας για τη συμμετοχή του στην Εξεταστική Επιτροπή, διαβεβαιώνοντας τον ότι οι παρατηρήσεις του θα ληφθούν σοβαρά υπ' όψιν και θα ενσωματωθούν στο τελικό κείμενο.

Το Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Στερεάς Ελλάδος και τους διδάσκοντες για το θετικό ακαδημαϊκό κλίμα και την άψογη συνεργασία μας.

Τη διδακτορική φοιτήτρια Παναγιώτα Κοντού, για την πολύτιμη συμβολή της στην ανάπτυξη τμήματος του κώδικα στο πλαίσιο της παρούσας εργασίας.

Τον Δήμο Αχαρνών για την οικονομική υποστήριξη που μου παρείχε, δεδομένου ότι κατά τη διάρκεια των σπουδών μου ήμουν υπότροφος του Κληροδοτήματος «Ροδία Στριφτού».

Επιπλέον, θα ήθελα να απευθύνω τις ευχαριστίες μου στην οικογένειά μου, η οποία με στήριξε ποικιλοτρόπως, τόσο ηθικά όσο και υλικά, καθ' όλη τη διάρκεια της μέχρι τώρα ζωής μου.

Τέλος, τους φίλους μου και ιδιαίτερα τις συμφοιτήτριες μου Παπαθανασίου Κωνσταντίνα και Παπαθεοδοσίου Μερσίνη καθώς και τον αρραβωνιαστικό μου Καμμένο Παναγιώτη τόσο για την εποικοδομητική βοήθεια που μου παρείχαν όσο και για την ψυχολογική υποστήριξη τους.

Ειρήνη Καλαθά
Λαμία, Ιούλιος 2012

Περιεχόμενα

Πρόλογος

Ένα από τα συχνά προβλήματα που αντιμετωπίζουν οι (βιο)στατιστικοί, οι οικονομολόγοι, οι ψυχολόγοι, κ.ά., οι οποίοι ασχολούνται με τη στατιστική ανάλυση, και ειδικότερα με την πολυμεταβλητή ανάλυση, είναι η παρουσία ελλιπών στοιχείων στα σύνολα δεδομένων που μελετούν- δηλαδή παρατηρήσεις και μετρήσεις οι οποίες θα έπρεπε να είχαν καταγραφεί αλλά για κάποιο λόγο δεν καταγράφτηκαν.

Η απουσία αυτών των στοιχείων προκαλεί κωλύματα σε διάφορες πολυμεταβλητές τεχνικές ανάλυσης δεδομένων, όπως για παράδειγμα στον υπολογισμό των βασικών στατιστικών πινάκων, στην παλινδρόμηση και στην ανάλυση κατά παράγοντες (factor analysis) [15].

Στην παρούσα πτυχιακή μελετάται η διαταραχή ιδιοτιμών συμμετρικών πινάκων η οποία συνιστά ένα τρόπο παράκαμψης των πρακτικών δυσχερειών που περιγράφθηκαν παραπάνω. Ειδικότερα, σκοπός της εν λόγω πτυχιακής είναι η ανασκόπηση της υπάρχουσας βιβλιογραφίας-αρθρογραφίας αναφορικά με τις μεθόδους υπολογισμού συμμετρικών πινάκων οι οποίοι προέρχονται από σύνολα δεδομένων με ελλιπή δεδομένα και τις μεθόδους διαταραχής ιδιοτιμών συμμετρικών πινάκων καθώς και η σύγκριση των μεθόδων διαταραχής.

Κάθε κεφάλαιο αυτής της πτυχιακής εργασίας αριθμείται και υποδιαιρείται σε ενότητες οι οποίες αριθμούνται με δυο αριθμούς, ενώ μερικά αριθμούνται με τρεις αριθμούς. Ο πρώτος αριθμός αναφέρεται στο κεφάλαιο, ο δεύτερος στην ενότητα και ο τρίτος, όπου υπάρχει, στην υποδιαίρεσή της. Επίσης, οι ορισμοί, τα θεωρήματα, οι προτάσεις και τα παραδείγματα αριθμούνται με δύο αριθμούς, από τους οποίους ο πρώτος αντιστοιχεί στο κεφάλαιο και ο δεύτερος στη σειρά εμφάνισης της ενότητας.

Στο πρώτο κεφάλαιο παρουσιάζονται οι βασικές έννοιες της θεωρίας πινάκων, όπως για παράδειγμα, ιδιοτιμές και ιδιοδιανύσματα καθώς και οι σημαντικότερες ιδιότητες τους. Επιπλέον, γίνεται αναφορά στους ορισμούς και στις ιδιότητες για τις νόρμες πίνακα και λαμβάνει χώρα μια εκτενής εισαγωγή στην πολυμεταβλητή στατιστική. Στο τέλος του κεφαλαίου παρουσιάζεται το μοντέλο της γραμμικής παλινδρόμησης.

Το δεύτερο κεφάλαιο επικεντρώνεται σε ένα θέμα το οποίο τις τελευταίες δεκαετίες έχει απασχολήσει πολλούς ερευνητές [1, 2, 3, 5, 9, 13, 20, 23, 26] από διάφορους επιστημονικούς κλάδους. Συγκεκριμένα, ασχολείται με την εκτίμηση στατιστικών πινάκων οι οποίοι προέρχονται από σύνολα δεδομένων τα οποία εκτός από ορισμένες πλήρεις εγγραφές περιέχουν και ελλιπή δεδομένα (missing data). Οι μέθοδοι που περιγράφονται και οι οποίες συμβάλλουν στην εκτίμηση των στατιστικών πινάκων είναι η μέθοδος διανυσμάτων παρατηρήσεων με ελλιπή δεδομένα (listwise deletion) [13, 20, 22, 23] , η μέθοδος EM (Expectation–maximization algorithm) [2, 5, 9, 15] και η μέθοδος διαγραφής ανά ζεύγη (pairwise deletion) [3, 13, 26].

Στο τρίτο κεφάλαιο, παρατίθενται μέθοδοι διαταραχής ιδιοτιμών συμμετρικών πινάκων προκειμένου να υπολογιστούν οι πλησιέστεροι προσεγγιστικά θετικά ημιορισμένοι πίνακες και να παρακαμφθούν τα πρακτικά κωλύματα που προκαλεί η αοριστία των πινάκων. Ειδικότερα, οι μέθοδοι διαταραχής ιδιοτιμών συμμετρικού πίνακα οι οποίες θα παρουσιαστούν στην παρούσα ενότητα αριθμούνται σε τρεις και είναι η μέθοδος μερικής φασματικής ανάλυσης, η μέθοδος της μεταφοράς και η μέθοδος μέσω πίνακα συσχέτισης. Επιπλέον, στο κεφάλαιο αυτό λαμβάνει χώρα σύγκριση των μεθόδων διαταραχής με τη χρήση της νόρμας Frobenius, και ενός κριτηρίου διαφοράς με τη χρήση απόλυτων τιμών καθώς και γραμμική παλινδρόμηση. Η σύγκριση των μεθόδων συμβάλλει στην εξαγωγή συμπερασμάτων αναφορικά με το ποια μέθοδος είναι βέλτιστη.

Στο τέλος της πτυχιακής εργασίας υπάρχουν τα συμπεράσματα, βιβλιογραφία με τα σχετικά συγγράμματα, τα οποία αναφέρονται στο κείμενο και πολλά από αυτά είναι χρήσιμα για περαιτέρω μελέτη των προαναφερόμενων εννοιών και εμβάθυνση, καθώς και τρία παραρτήματα. Στο πρώτο δίνεται ο κώδικας σε Stata που αναπτύχθηκε στην παρούσα εργασία για τον υπολογισμό ενός πίνακα συνδιακύμανσης, ο οποίος προέρχεται από ένα σύνολο δεδομένων με ελλιπή δεδομένα.

Συγκεκριμένα, υλοποιήθηκε η μέθοδος παρατηρήσεων με ελλιπή δεδομένα καθώς και η μέθοδος διαγραφής ανά ζεύγη. Στο δεύτερο παράρτημα παρατίθεται κώδικας σε Stata ο οποίος υλοποιεί τις μεθόδους διαταραχής ιδιοτιμών ενός αόριστου πίνακα συνδιακύμανσης. Ειδικότερα, οι διορθωτικές μέθοδοι διαταραχής οι οποίες υλοποιούνται είναι η μέθοδος ειδικής φασματικής ανάλυσης, της μεταφοράς και η μέθοδος μέσω πίνακα συσχέτισης. Τέλος, στο τρίτο παράρτημα παρουσιάζονται τα σύνολα δεδομένων που χρησιμοποιήθηκαν στην παρούσα πτυχιακή εργασία κατά την υλοποίηση του κώδικα Stata και κατά την εξαγωγή συμπερασμάτων.

1^o ΚΕΦΑΛΑΙΟ

ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΠΙΝΑΚΩΝ ΚΑΙ ΠΟΛΥΜΕΤΑΒΛΗΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

1.1 Άλγεβρα πινάκων

1.1.1 Βασικοί ορισμοί

Ορισμός 1.1

Μια διάταξη των mn στοιχείων (αριθμών) a_{ij} του συνόλου $\mathbb{F}$ (σύνολο $\mathbb{R}$ των πραγματικών ή σύνολο $\mathbb{C}$ των μιγαδικών αριθμών) σε σχήμα ορθογωνίου παραλληλογράμμου με στοιχεία από το $\mathbb{F}$ ονομάζεται πίνακας A με στοιχεία από το $\mathbb{F}$ ο οποίος σημειώνεται ως εξής:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Ο παραπάνω πίνακας A είναι ένας πίνακας $m \times n$, όπου m είναι το πλήθος των γραμμών του και n το πλήθος των στηλών του και αναφερόμαστε στο πίνακα χρησιμοποιώντας το **μέγεθος** ή τον **τύπο** του πίνακα.

Οι αριθμοί a_{ij} ονομάζονται **στοιχεία** του πίνακα, ο δε δείκτης ij αναφέρεται στις γραμμές / στήλες του A, ο i στις γραμμές και ο j στις στήλες.

Επιπροσθέτως, ο πίνακας A συμβολίζεται (a_{ij}), $i=1,...,m, j=1,...,n$ και με $M_{m\times n}(\mathbb{F})$ συμβολίζεται το σύνολο όλων των $m\times n$ πινάκων με στοιχεία από το $\mathbb{F}$. Δύο ή περισσότεροι πίνακες που έχουν ίδιο αριθμό γραμμών και στηλών χαρακτηρίζονται ως πίνακες **ιδίου τύπου**.

Ένας πίνακας, που έχει τον ίδιο αριθμό γραμμών και στηλών $(m=n)$, ονομάζεται **τετραγωνικός**. Το σύνολο όλων των $n \times n$ πινάκων με στοιχεία από το $\mathbb{F}$ συμβολίζεται με $M_n(\mathbb{F})$.

Στη συνέχεια αναφέρονται ορισμένες κατηγορίες πινάκων, που παρουσιάζουν ιδιαίτερο ενδιαφέρον και θα συναντήσουμε στην παρούσα εργασία.

Ορισμός 1.2

i. Ένας $1 \times n$ πίνακας λέγεται και **πίνακας-γραμμή**, ενώ ένας $m \times 1$ πίνακας λέγεται και **πίνακας-στήλη** ή **διάνυσμα**.

ii. Εάν όλα τα στοιχεία ενός $m \times n$ πίνακα είναι ίσα με μηδέν, ο πίνακας αυτός ονομάζεται **μηδενικός** και συμβολίζεται με $0_{m \times n}$ ή απλά με 0.

iii. Έστω $A=(a_{ij})$ ένας $m \times n$ πίνακας. Ο $n \times m$ πίνακας (a_{ji}) ονομάζεται **ανάστροφος** του A και συμβολίζεται με A^T και $n \times m$ πίνακας $(\overline{a}_{ji})$ ονομάζεται **αναστροφοσυζυγής** του A και συμβολίζεται $A^* = \overline{A}^T$.

iv. Έστω $A=(a_{ij})$ ένας $m \times n$ πίνακας. Ο $m \times n$ πίνακας $(\overline{a}_{ij})$, ο οποίος έχει ως στοιχεία του τα συζυγή στοιχεία του A, ονομάζεται **συζυγής** του A και συμβολίζεται $\overline{A}$.

v. Ένας πίνακας $A \in M_n(\mathbb{F})$ λέγεται **διαγώνιος**, αν για κάθε $i \neq j$ ισχύει $a_{ij}=0$, δηλαδή αν κάθε στοιχείο που δε βρίσκεται στη διαγώνιο, είναι ίσο με μηδέν. Οι πίνακες αυτοί συμβολίζονται και ως $A = diag(a_{11}, a_{22}, ..., a_{nn})$.

Ειδικότερα, ο διαγώνιος πίνακας που έχει όλα τα διαγώνια στοιχεία ίσα με 1 ονομάζεται **μοναδιαίος** και συμβολίζεται $\boldsymbol{I}_n = diag(1,1,...,1)$.

vi. Ένας πίνακας $A \in M_n(\mathbb{F})$ λέγεται **άνω τριγωνικός**, αν για κάθε $i > j$ ισχύει $a_{ij}=0$, δηλαδή αν τα στοιχεία που βρίσκονται κάτω από την κύρια διαγώνιο είναι μηδέν.

Αντίθετα ένας πίνακας $A \in M_n(\mathbb{F})$ λέγεται **κάτω τριγωνικός**, αν για κάθε $i < j$ ισχύει $a_{ij}=0$, δηλαδή αν τα στοιχεία που βρίσκονται πάνω από την κύρια διαγώνιο είναι μηδέν.

Ορισμός 1.3

Δύο πίνακες $A, B \in M_n(\mathbb{F})$ ονομάζονται **όμοιοι** αν για αυτούς υπάρχει αντιστρέψιμος[1] πίνακας $P \in M_n(\mathbb{F})$ τέτοιος ώστε να ισχύει $A = PBP^{-1}$.

Ο πίνακας P ονομάζεται πίνακας ομοιότητας.

Ορισμός 1.4

Ένας πίνακας $A = (a_{ij}) \in M_n(\mathbb{R})$ λέγεται **συμμετρικός**, αν για κάθε i, j ισχύει $a_{ij} = a_{ji}$. Ισοδύναμα, ένας πίνακας A είναι συμμετρικός αν και μόνο αν ισχύει $A = A^T$.

Για παράδειγμα, ο πίνακας $A = \begin{bmatrix} 1 & \sqrt{3} & -1 \\ \sqrt{3} & 0 & 2 \\ -1 & 2 & 3 \end{bmatrix}$ είναι συμμετρικός.

Ορισμός 1.5

Έστω $A \in M_n(\mathbb{R})$ συμμετρικός πίνακας. Αν για κάθε μη μηδενικό $\boldsymbol{x} \in M_{n\times 1}(\mathbb{R})$ ισχύει

$$\boldsymbol{x}^T A \boldsymbol{x} > 0,$$

τότε ο πίνακας A ονομάζεται **θετικά ορισμένος** (positive definite), ενώ όταν ισχύει $\boldsymbol{x}^T A\boldsymbol{x} \geq 0$ ονομάζεται **θετικά ημιορισμένος πίνακας** (positive semidefinite matrix) ή **μη-αρνητικά ορισμένος πίνακας** (non-negative definite matrix).

Ορισμός 1.6

Έστω $A \in M_n(\mathbb{R})$ συμμετρικός πίνακας. Αν για κάθε μη μηδενικό $\boldsymbol{x} \in M_{n\times 1}(\mathbb{R})$ ισχύει

$$\boldsymbol{x}^T A \boldsymbol{x} < 0,$$

τότε ο πίνακας A ονομάζεται **αρνητικά ορισμένος** (negative definite), ενώ όταν ισχύει $\boldsymbol{x}^T A\boldsymbol{x} \leq 0$ ονομάζεται **αρνητικά ημιορισμένος πίνακας** (negative semidefinite matrix).

---

[1] Βλέπε, Ορισμό 1.8 και την ικανή και αναγκαία συνθήκη στην Πρόταση 1.1, ιδιότητα (ii).

Αν ένας πίνακας δεν είναι θετικά (αρνητικά) ορισμένος ονομάζεται **αόριστος**.

Ορισμός 1.7

Αν για έναν πίνακα $A \in M_n(\mathbb{C})$ ισχύει

$$A^* A = AA^* = I_n,$$

ο πίνακας ονομάζεται **ορθομοναδιαίος**,

ενώ αν $A \in M_n(\mathbb{R})$ και ισχύει

$$A^T A = AA^T = I_n,$$

ο πίνακας ονομάζεται **ορθογώνιος**.

Ορισμός 1.8

Ένας τετραγωνικός πίνακας $A \in M_n(\mathbb{F})$ ονομάζεται **αντιστρέψιμος** αν ισχύει

$$AB = BA = I_n,$$

όπου $B \in M_n(\mathbb{F})$ και ονομάζεται **αντίστροφος** του A.

Ορισμός 1.9

Έστω δυο πίνακες $A, B \in M_{m\times n}(\mathbb{F})$ με $A = a_{ij}$, $B = b_{ij}$.

Το **Hadamard γινόμενο** των πινάκων A, B συμβολίζεται $A \circ B$, και είναι ο $m\times n$ πίνακας $A \circ B = c_{ij}$ όπου $c_{ij} = a_{ij}b_{ij}$, για όλα τα $1 \le i \le m$, $1 \le j \le n$.

1.1.2 Χαρακτηριστικά μεγέθη

Ορισμός 1.10

Ορίζουσα ενός τετραγωνικού πίνακα $A \in M_n(\mathbb{F})$ είναι μια απεικόνιση

$$\det : M_n(\mathbb{F}) \to \mathbb{F} : A \to \det A$$

για την οποία ισχύει

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}),$$

για οποιοδήποτε $j = 1, 2, ..., n$, όπου A_{ij} είναι $(n-1)\times(n-1)$ πίνακας, ο οποίος προκύπτει από τον πίνακα A, αν διαγράψουμε τα στοιχεία της i-γραμμής και της j-στήλης.

Η ορίζουσα του $A \in M_n(\mathbb{F})$ συμβολίζεται $|A|$. Ειδικά όταν ξέρουμε τα στοιχεία του πίνακα A, γράφουμε $|A| = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$, $\det(A)$ ή απλούστερα $\det A$.

Προφανώς από τον παραπάνω Ορισμό 1.10 είναι φανερό ότι η ορίζουσα ενός τετραγωνικού πίνακα είναι αριθμός πραγματικός ή μιγαδικός.

Για παράδειγμα, ο πίνακας $A = \begin{bmatrix} -1 & 2 & -1 \\ 2 & 1 & 0 \\ 1 & 4 & -2 \end{bmatrix}$ έχει ορίζουσα, που υπολογίζεται από τύπο στον Ορισμό 1.10 και είναι:

$$\det \begin{bmatrix} -1 & 2 & -1 \\ 2 & 1 & 0 \\ 1 & 4 & -2 \end{bmatrix} = (-1)^{1+3}(-1)\begin{vmatrix} 2 & 1 \\ 1 & 4 \end{vmatrix} + (-1)^{2+3}\cdot 0 \cdot \begin{vmatrix} -1 & 2 \\ 1 & 4 \end{vmatrix} + (-1)^{3+3}(-2)\begin{vmatrix} -1 & 2 \\ 2 & 1 \end{vmatrix}$$

$$= -(2\cdot 4 - 1\cdot 1) + 0 + (-2)[(-1)\cdot 1 - 2\cdot 2] = 3$$ ◊

Οι σημαντικότερες **ιδιότητες** των οριζουσών, η απόδειξη των οποίων δίνεται στα βιβλία [4, 7, 16], είναι οι ακόλουθες:

Πρόταση 1.1

i. *Για κάθε πίνακα* $A \in M_n(\mathbb{F})$ *οι ιδιότητες που αναφέρονται στις γραμμές της ορίζουσας του πίνακα, ισχύουν και για τις στήλες του.*

ii. *Ένας πίνακας* $A \in M_n(\mathbb{F})$ *είναι αντιστρέψιμος αν και μόνο αν* $\det A \neq 0$.

iii. *Αν ένας πίνακας* $A \in M_n(\mathbb{F})$ *έχει μηδενική γραμμή, τότε* $\det A = 0$.

iv. *Αν ένας πίνακας* $A \in M_n(\mathbb{F})$ *έχει δύο γραμμές ανάλογες, τότε* $\det A = 0$.

v. *Για κάθε άνω (κάτω) τριγωνικό πίνακα* $A \in M_n(\mathbb{F})$ *ισχύει* $\det A = a_{11}a_{22}...a_{nn}$, *όπου* a_{ii}, $i = 1, 2, ..., n$, *είναι τα διαγώνια στοιχεία του πίνακα.*

vi. *Ένας διαγώνιος έχει ορίζουσα που ισούται με το γινόμενο των διαγώνιων στοιχείων του. Η ορίζουσα του μοναδιαίου πίνακα ισούται με* 1.

vii. *Για κάθε πίνακα* $A \in M_n(\mathbb{F})$ *ισχύει* $\det(A^T) = \det A$, $\det(\overline{A}) = \overline{\det A}$ *και* $\det(A^*) = \overline{\det A}$.

viii. *Αν* $A, B \in M_n(\mathbb{F})$, *τότε* $\det(AB) = \det A \cdot \det B$.

ix. *Για κάθε πίνακα* $A \in M_n(\mathbb{F})$ *ισχύει* $\det(A^k) = (\det A)^k$.

x. *Για έναν πίνακα* $A \in M_n(\mathbb{F})$ *και* $\lambda \in \mathbb{F}$ *ισχύει* $\det(\lambda A) = \lambda^n (\det A)$.

Ορισμός 1.11

Έστω ο τετραγωνικός πίνακας $A \in M_n(\mathbb{F})$ και ένα διάνυσμα $\boldsymbol{x} \in M_{n\times 1}(\mathbb{F})$ τέτοιο ώστε να ισχύει

$$A\boldsymbol{x} = \lambda \boldsymbol{x}, \text{ με } \boldsymbol{x} \neq \boldsymbol{0} \tag{1.1}$$

για κάποιο $\lambda \in \mathbb{F}$. Η τιμή του λ ονομάζεται **ιδιοτιμή** (eigenvalue) του πίνακα A και το $\boldsymbol{x}$ ονομάζεται **ιδιοδιάνυσμα** του A, αντίστοιχο της ιδιοτιμής λ.

Οι ιδιοτιμές και τα ιδιοδιανύσματα ονομάζονται **χαρακτηριστικά ποσά** ή **ιδιοποσά** του πίνακα A.

Το σύνολο των ιδιοτιμών του πίνακα A συμβολίζεται $\sigma(A)$ και ονομάζεται **φάσμα** (spectrum) του πίνακα.

Από όλες τις ιδιοτιμές του πίνακα, αυτή που έχει τη μεγαλύτερη απόλυτη τιμή ονομάζεται **φασματική ακτίνα** (spectral radius) και συμβολίζεται $\rho(A)$, δηλαδή

$$\rho(A) = \max\{|\lambda| : \ \lambda \in \sigma(A)\}. \tag{1.2}$$

Για τον υπολογισμό όλων των ιδιοτιμών $\lambda \in \mathbb{F}$ ενός τετραγωνικού πίνακα A, πρέπει να ισχύει η διανυσματική εξίσωση στην (1.1), η οποία οδηγεί σε ένα γραμμικό σύστημα, που έχει μη μηδενική λύση αν και μόνο αν

$$\det(A - \lambda I) = 0$$

και αναλυτικότερα

$$\det \begin{bmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{bmatrix} = 0 \tag{1.3}$$

Συνεπώς, όλες οι ιδιοτιμές $\lambda \in \mathbb{F}$ του πίνακα A πρέπει να ικανοποιούν την εξίσωση (1.3), η οποία ονομάζεται **χαρακτηριστική εξίσωση** του τετραγωνικού πίνακα A.

Αναπτύσσοντας την ορίζουσα στο αριστερό μέρος της χαρακτηριστικής εξίσωσης (1.3) καταλήγουμε σε ένα πολυώνυμο n-οστού βαθμού, το οποίο γράφεται

$$\chi_A(\lambda) = (-1)^n \det(A - \lambda I) = \det(\lambda I - A) = \lambda^n + b_{n-1}\lambda^{n-1} + \dots + b_1\lambda + b_0 \,,$$

και ονομάζεται **χαρακτηριστικό πολυώνυμο** του A με συντελεστές $b_j \in \mathbb{F}$, $j = 0,1,...,n-1$.

Το χαρακτηριστικό πολυώνυμο μπορεί να αναλυθεί πάντοτε σε πρωτοβάθμιους παράγοντες στο $\mathbb{C}$ και να γραφεί στη μορφή

$$\chi_A(\lambda) = (\lambda - \lambda_1)^{\nu_1}(\lambda - \lambda_2)^{\nu_2} \cdots (\lambda - \lambda_i)^{\nu_i} \,,$$

όπου $\lambda_1, \lambda_2, ..., \lambda_i$, είναι όλες οι διαφορετικές (διακεκριμένες) ρίζες του $\chi_A(\lambda)$ στο $\mathbb{C}$ και $\nu_1, \nu_2, ..., \nu_i$ η πολλαπλότητα κάθε ρίζας, η οποία ονομάζεται **αλγεβρική πολλαπλότητα**.

Είναι φανερό ότι, για έναν πίνακα $A \in M_n(\mathbb{F})$ ισχύει

$$\nu_1 + \nu_2 + \dots + \nu_i = n \,.$$

Αν για κάποιο i ισχύει $\nu_i = 1$, η ιδιοτιμή χαρακτηρίζεται **διακεκριμένη** ή **απλή**, διαφορετικά ονομάζεται **πολλαπλή**.

Η ισότητα (1.1) γράφεται ισοδύναμα :

$$A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \Leftrightarrow (A - \lambda I)\mathbf{x} = \mathbf{0} \,.$$

Αν γνωρίζουμε την ιδιοτιμή λ είναι εύκολο να υπολογίσουμε το αντίστοιχο μη μηδενικό ιδιοδιάνυσμα, λύνοντας το γραμμικό σύστημα που προκύπτει από την εξίσωση :

$$(A-\lambda I)\boldsymbol{x}=\boldsymbol{0}\text{, για } \boldsymbol{x}\neq\boldsymbol{0}.$$

Αντικαθιστώντας κάθε διακεκριμένη ιδιοτιμή, $\lambda=\lambda_i$, στο σύστημα (1.1) ή στο παραπάνω ισοδύναμο του, παίρνουμε ως γενική λύση του ομογενούς συστήματος $(A-\lambda I)\boldsymbol{x}=\boldsymbol{0}$ ένα μη κενό υποσύνολο του $M_{n\times 1}(\mathbb{F})$, που ονομάζεται **ιδιόχωρος** και συμβολίζεται με $V(\lambda_i)$, δηλαδή του

$$V(\lambda_i)=\{x\in M_{n\times 1}(\mathbb{F}):(A-\lambda_i I)x=0\}.$$

Τα μη μηδενικά στοιχεία του $V(\lambda_i)$ είναι τα ιδιοδιανύσματα που αντιστοιχούν στην ιδιοτιμή λ_i.

Στη συνέχεια, διατυπώνεται μια βασική ιδιότητα που αναφέρεται στην ιδιοτιμή ενός τετραγωνικού πίνακα A.

Πρόταση 1.2

Έστω $A\in M_n(\mathbb{F})$ *με ιδιοτιμές* λ_i, $i=1,2,\ldots,n$ *και* $\tilde{A}$ *ένας πίνακας τέτοιος ώστε*

$$\tilde{A}=A+kI_n$$

για οποιοδήποτε αριθμό k. *Οι ιδιοτιμές του πίνακα* $\tilde{A}$ *είναι* λ_i+k *για κάθε* $i=1,2,\ldots,n$.

Απόδειξη : Θεωρώντας ότι λ_i είναι κάποια ιδιοτιμή του A και $\boldsymbol{x}_i$ το αντίστοιχο ιδιοδιάνυσμά της, χρησιμοποιώντας τον Ορισμό 1.11 και τη σχέση (1.1) μπορούμε να γράψουμε $A\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i$, οπότε για τον πίνακα $\tilde{A}$ και το διάνυσμα $\boldsymbol{x}_i$ έχουμε

$$\tilde{A}\boldsymbol{x}_i=(A+kI_n)\boldsymbol{x}_i=A\boldsymbol{x}_i+kI_n\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i+k\boldsymbol{x}_i=(\lambda_i+k)\boldsymbol{x}_i$$

το οποίο επαληθεύει τον Ορισμό 1.11 για τον πίνακα $\tilde{A}$. □

Για τα χαρακτηριστικά ποσά ενός τετραγωνικού πίνακα $A\in M_n(\mathbb{F})$ οι βασικότερες **ιδιότητες** διατυπώνονται στην επόμενη πρόταση, η απόδειξη των οποίων μπορεί να αναζητηθεί σε οποιοδήποτε σύγγραμμα Γραμμικής Άλγεβρας, που αναφέρεται στη βιβλιογραφία, [4, 7, 12, 14, 16, 21].

Πρόταση 1.3

Έστω $A \in M_n(\mathbb{F})$ *και* $\lambda_1, \lambda_2, ..., \lambda_n$ *είναι οι ιδιοτιμές*[2] *του.*

i. *Οι ιδιοτιμές ενός άνω (κάτω) τριγωνικού πίνακα και ενός διαγώνιου πίνακα είναι τα διαγώνια στοιχεία τους.*

ii. i. $\det A = \lambda_1 \cdot \lambda_2 \cdot \cdots \cdot \lambda_n$.

   ii. *Ο πίνακας* $A \in M_n(\mathbb{F})$ *είναι αντιστρέψιμος, αν και μόνο αν δεν έχει το μηδέν ως ιδιοτιμή.*

   iii. $trA = \lambda_1 + \lambda_2 + \cdots + \lambda_n$, *δηλαδή το άθροισμα όλων των ιδιοτιμών του* A *ισούται με το ίχνος του,* trA.

iii. *Για κάθε* $A \in M_n(\mathbb{F})$ *ισχύει* $\sigma(A) = \sigma(A^T)$.

iv. *Αν* λ, $\mathbf{x}$ *είναι χαρακτηριστικά ποσά του* $A \in M_n(\mathbb{F})$, *τότε για τα χαρακτηριστικά ποσά του πίνακα* A^k *ισχύει* $A^k \mathbf{x} = \lambda^k \mathbf{x}$, *όπου* $k \in \mathbb{N}$.

v. *Αν* λ, $\mathbf{x}$ *είναι τα χαρακτηριστικά ποσά ενός αντιστρέψιμου πίνακα* $A \in M_n(\mathbb{F})$, *τότε τα χαρακτηριστικά ποσά του πίνακα* A^{-k} *είναι* λ^{-k} *και* $\mathbf{x}$, *όπου* $k \in \mathbb{N}$.

vi. *Οι όμοιοι*[3] *πίνακες έχουν τις ίδιες ιδιοτιμές.*

vii. *Οι ιδιοτιμές ενός συμμετρικού*[4] *πίνακα* $A \in M_n(\mathbb{F})$ *είναι πραγματικοί αριθμοί.*

viii. *Ένας συμμετρικός πίνακας* $A \in M_n(\mathbb{R})$ *είναι θετικά ορισμένος πίνακας* αν και μόνο αν *οι ιδιοτιμές του είναι θετικοί πραγματικοί αριθμοί* και είναι *θετικά ημιορισμένος αν και μόνο αν μεταξύ των θετικών ιδιοτιμών του υπάρχει τουλάχιστον μια ίση με μηδέν.*

ix. *Ένας συμμετρικός πίνακας* $A \in M_n(\mathbb{R})$ *είναι αρνητικά ορισμένος πίνακας* αν και μόνο αν *οι ιδιοτιμές του είναι αρνητικοί πραγματικοί αριθμοί* και είναι *αρνητικά ημιορισμένος αν και μόνο αν μεταξύ των αρνητικών ιδιοτιμών του υπάρχει τουλάχιστον μια ίση με μηδέν.*

---

[2] Οι ιδιοτιμές δεν είναι απαραίτητα διακεκριμένες.

[3] Βλέπε Ορισμό 1.3.

[4] Βλέπε Ορισμό 1.4.

Χρήσιμος χαρακτηρισμός των ιδιοδιανυσμάτων είναι αυτός που διατυπώνεται στον ακόλουθο ορισμό.

Ορισμός 1.12

Έστω V ένας $\mathbb{F}$-διανυσματικός χώρος. Τα διανύσματα $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n \in M_{n\times 1}(\mathbb{F})$ του V ονομάζονται **γραμμικά ανεξάρτητα** αν και μόνο αν η διανυσματική εξίσωση

$$a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2 + \cdots + a_n\boldsymbol{u}_n = \mathbf{0},\ \mu\varepsilon\ a_1, a_2, \ldots, a_n \in \mathbb{F} \qquad (1.4)$$

έχει μοναδική λύση την τετριμμένη, $a_1 = a_2 = \ldots = a_n = 0$.

Αν η (1.4) έχει και άλλες λύσεις εκτός της μηδενικής, τα διανύσματα $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ ονομάζονται **γραμμικά εξαρτημένα**.

Ορισμός 1.13

Ένας πίνακας $A \in M_n(\mathbb{F})$ ονομάζεται **διαγωνοποιήσιμος** όταν είναι όμοιος[5] με ένα διαγώνιο πίνακα Δ που έχει διαγώνια στοιχεία τις ιδιοτιμές λ_i του A, δηλαδή ισχύει

$$A = P\Delta P^{-1},$$

όπου

$$\Delta = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

και P είναι ένας τετραγωνικός πίνακας με στήλες τα ιδιοδιανύσματα $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ του A,

$$P = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix}.$$

Αποδεικνύεται ότι ένας πίνακας $A \in M_n(\mathbb{F})$ είναι διαγωνοποιήσιμος, αν και μόνο αν ο πίνακας A έχει n γραμμικά ανεξάρτητα ιδιοδιανύσματα[6], η απόδειξη της πρότασης βρίσκεται στο [7]. Σημειώνεται ότι σε διακεκριμένες ιδιοτιμές αντιστοιχούν γραμμικά ανεξάρτητα ιδιοδιανύσματα, συνεπώς αν ο πίνακας $A \in M_n(\mathbb{F})$ έχει n διακεκριμένες ιδιοτιμές, τότε είναι διαγωνοποιήσιμος. Σε περίπτωση όπου οι ιδιοτιμές

---

[5] Βλέπε Ορισμό 1.3.

[6] Βλέπε Ορισμό 1.12.

έχουν αλγεβρική πολλαπλότητα μεγαλύτερη από 1, **δεν** υπάρχει αντίστοιχη πρόταση ανεξαρτησίας των ιδιοδιανυσμάτων, που αντιστοιχούν στην πολλαπλή ιδιοτιμή, γεγονός που δεν εγγυάται την ύπαρξη τετραγωνικού και αντιστρέψιμου πίνακα P που να διαγωνοποιεί τον A.

Στην ειδική περίπτωση όπου ο πίνακας A είναι συμμετρικός το **φασματικό θεώρημα** εξασφαλίζει την ύπαρξη καθώς και την αντιστρεψιμότητα του P, ανεξάρτητα από την πολλαπλότητα των πραγματικών[7] ιδιοτιμών του A.

Θεώρημα 1.1 (φασματικό θεώρημα)

Κάθε συμμετρικός[8] *πίνακας* $A \in M_n(\mathbb{R})$ *είναι ορθογώνια όμοιος με πραγματικό διαγώνιο πίνακα, δηλαδή*

$$A = U\Delta U^T \tag{1.5}$$

όπου ο Δ *είναι διαγώνιος πίνακας, που έχει διαγώνια στοιχεία τις ιδιοτιμές του* A *και* U *είναι ορθογώνιος*[9] *πίνακας, που έχει ως στήλες τα αντίστοιχα ορθογώνια ιδιοδιανύσματα των ιδιοτιμών με τη σειρά που παρουσιάστηκαν στα διαγώνια στοιχεία του* Δ.

Σύμφωνα με το Θεώρημα 1.1 ένας συμμετρικός πίνακας A διαγωνοποιείται πάντα ανεξάρτητα από την πολλαπλότητα που παρουσιάζουν οι ιδιοτιμές. Για έναν τέτοιον πίνακα μπορεί να διατυπωθεί μια διαφορετική ανάλυση από την (1.5) του συμμετρικού πίνακα, η οποία διατυπώνεται στο επόμενο θεώρημα. Η ανάλυση αυτή σχετίζεται με τις ιδιοτιμές και τα ιδιοδιανύσματα του A και αποδεικνύεται [14, 21].

Θεώρημα 1.2 (φασματική ανάλυση)

Για κάθε συμμετρικό πίνακα $A \in M_n(\mathbb{R})$ *η φασματική ανάλυση δίνεται από την ισότητα:*

$$A = \lambda_1 \boldsymbol{u}_1 \boldsymbol{u}_1^T + \lambda_2 \boldsymbol{u}_2 \boldsymbol{u}_2^T + \cdots + \lambda_n \boldsymbol{u}_n \boldsymbol{u}_n^T \tag{1.6}$$

όπου $\lambda_1, \lambda_2, \ldots, \lambda_n$ *είναι οι ιδιοτιμές του πίνακα* A *και τα* $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ *είναι τα αντίστοιχα ορθογώνια ιδιοδιανύσματά του.*

---

[7] Βλέπε ιδιότητα vii, Πρόταση 1.3.

[8] Βλέπε Ορισμό 1.4.

[9] Βλέπε Ορισμό 1.7.

Παράδειγμα 1.1

Έστω ο συμμετρικός πίνακας

$$A=\begin{bmatrix}5 & 2\\ 2 & 8\end{bmatrix}.$$

Με τη χρήση της (1.3) προκύπτει το χαρακτηριστικό πολυώνυμο, που είναι

$$\chi_A(\lambda)=\det(\lambda I-A)=\det\begin{bmatrix}\lambda-5 & -2\\ -2 & \lambda-8\end{bmatrix}=(\lambda-5)(\lambda-8)-4=\lambda^2-13\lambda+36$$

Συνεπώς, οι ιδιοτιμές του A είναι οι ρίζες του παραπάνω χαρακτηριστικού πολυωνύμου, $\lambda_1=9, \lambda_2=4$.

Αναζητώντας τα αντίστοιχα ιδιοδιανύσματά τους $\boldsymbol{x}=\begin{bmatrix}\mathrm{x}_1\\ \mathrm{x}_2\end{bmatrix}$, από την (1.1) καταλήγουμε στο ομογενές σύστημα

$$\begin{aligned}(5-\lambda)\mathrm{x}_1+2\mathrm{x}_2&=0\\ 2\mathrm{x}_1+(8-\lambda)\mathrm{x}_2&=0\end{aligned}$$

Για $\lambda_1=9$, το αντίστοιχο ομογενές σύστημα δίνει

$$\left.\begin{aligned}-4\mathrm{x}_1+2\mathrm{x}_2=0\\ 2\mathrm{x}_1-\mathrm{x}_2=0\end{aligned}\right\}\Rightarrow \boldsymbol{x}\ =\begin{bmatrix}\mathrm{x}_1\\ 2\mathrm{x}_1\end{bmatrix}=\begin{bmatrix}1\\ 2\end{bmatrix}\mathrm{x}_1,\quad \mathrm{x}_1\in\mathbb{R}.$$

Άρα ο ιδιόχωρος $V(9)=\{\mathrm{x}_1\begin{bmatrix}1 & 2\end{bmatrix}^T, \mathrm{x}_1\in\mathbb{R}\}$ από όπου επιλέγουμε ως αντίστοιχο ιδιοδιάνυσμα της $\lambda_1=9$, $\boldsymbol{x}_1=\begin{bmatrix}1\\ 2\end{bmatrix}$. Με όμοιο τρόπο βρίσκουμε ότι, στην ιδιοτιμή $\lambda_2=4$ αντιστοιχεί ο ιδιόχωρος $V(9)=\{\mathrm{x}_2\begin{bmatrix}-2 & 1\end{bmatrix}^T, \mathrm{x}_2\in\mathbb{R}\}$ από όπου επιλέγουμε το ιδιοδιάνυσμα $\boldsymbol{x}_2=\begin{bmatrix}-2\\ 1\end{bmatrix}$.

Προκειμένου να επαληθεύσουμε το Θεώρημα 1.1 κάνουμε ορθογώνια τα ιδιοδιανύσματα $\boldsymbol{x}_1, \boldsymbol{x}_2$. Επειδή, το εσωτερικό τους γινόμενο είναι $\langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle=1(-2)-2(1)=0$, προφανώς τα $\boldsymbol{x}_1, \boldsymbol{x}_2$ είναι κάθετα.

Στη συνεχεία, υπολογίζουμε τα μέτρα των $\boldsymbol{x}_1, \boldsymbol{x}_2$, τα οποία είναι $\|\boldsymbol{x}_1\| = \|\boldsymbol{x}_2\| = \sqrt{5}$.

Συνεπώς, $\boldsymbol{u}_1 = \dfrac{\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|} = \begin{bmatrix} \dfrac{1}{\sqrt{5}} \\ \dfrac{2}{\sqrt{5}} \end{bmatrix}$ και $\boldsymbol{u}_2 = \dfrac{\boldsymbol{x}_2}{\|\boldsymbol{x}_2\|} = \begin{bmatrix} \dfrac{-2}{\sqrt{5}} \\ \dfrac{1}{\sqrt{5}} \end{bmatrix}$.

Άρα ο ορθογώνιος πίνακας U είναι:

$$U = \begin{bmatrix} \dfrac{1}{\sqrt{5}} & \dfrac{-2}{\sqrt{5}} \\ \dfrac{2}{\sqrt{5}} & \dfrac{1}{\sqrt{5}} \end{bmatrix}.$$

Χρησιμοποιώντας το διαγώνιο πίνακα των ιδιοτιμών

$$\Delta = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix}$$

η σχέση (1.5) επαληθεύεται μετά από τις ανάλογες πράξεις.

Αντικαθιστώντας τις ιδιοτιμές λ_1, λ_2 και τα αντίστοιχα ορθογώνια ιδιοδιανύσματα $\boldsymbol{u}_1, \boldsymbol{u}_2$ στη σχέση (1.6) παρατηρούμε ότι και αυτή επαληθεύεται:

$$\lambda_1 \boldsymbol{u}_1 \boldsymbol{u}_1^{\mathrm{T}} + \lambda_2 \boldsymbol{u}_2 \boldsymbol{u}_2^{\mathrm{T}} = 9 \begin{bmatrix} \dfrac{1}{\sqrt{5}} \\ \dfrac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \dfrac{1}{\sqrt{5}} & \dfrac{2}{\sqrt{5}} \end{bmatrix} + 4 \begin{bmatrix} \dfrac{-2}{\sqrt{5}} \\ \dfrac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \dfrac{-2}{\sqrt{5}} & \dfrac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 2 & 8 \end{bmatrix} = A \qquad \Diamond$$

1.2 Νόρμες

1.2.1 Νόρμες πινάκων

Ορισμός 1.14

Νόρμα ενός τετραγωνικού **πίνακα** $A \in M_n(\mathbb{F})$ είναι μια πραγματική απεικόνιση

$$\|\cdot\| : M_n(\mathbb{F}) \to [0, +\infty)$$

με τις εξής ιδιότητες:

i) $\|A\| \geq 0$

ii) $\|A\| = 0 \Leftrightarrow A = 0$

iii) $\|\lambda A\| = |\lambda| \, \|A\|$, για κάθε $\lambda \in \mathbb{F}$

iv) $\|A + B\| \leq \|A\| + \|B\|$, για κάθε $A, B \in M_n(\mathbb{F})$

v) $\|A \cdot B\| \leq \|A\| \cdot \|B\|$, για κάθε $A, B \in M_n(\mathbb{F})$

Για έναν πίνακα $A \in M_n(\mathbb{F})$ οι πιο γνωστές νόρμες [6] είναι οι ακόλουθες:

$$|||A|||_\infty = \max_{1 \leq i \leq n}\{r_i = \sum_{j=1}^{n} |a_{ij}|\} \text{ ή } ||A||_\Gamma \quad \text{(νόρμα γραμμής)} \tag{1.7}$$

$$|||A|||_1 = \max_{1 \leq j \leq n}\{c_i = \sum_{i=1}^{n} |a_{ij}|\} \text{ ή } ||A||_\Sigma \quad \text{(νόρμα στήλης)} \tag{1.8}$$

$$|||A|||_E = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}\right|^2} \quad \text{(Ευκλείδεια νόρμα)} \tag{1.9}$$

$$|||A|||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}\right|^2} = \sqrt{tr(AA^*)} \quad \text{(νόρμα Frobenius)} \tag{1.10}$$

$$|||A|||_2 = \sqrt{\rho(A^*A)} \quad \text{(φασματική νόρμα)} \tag{1.11}$$

όπου $\rho(A^*A)$ είναι η φασματική ακτίνα[10] του συμμετρικού πίνακα A^*A.

---

[10] Βλέπε Ορισμό 1.10, σχέση (1.2).

Σχόλιο 1.1

i. Η Ευκλείδεια νόρμα, η οποία δίνεται από τη σχέση (1.9), είναι ίδια με τη νόρμα Frobenius που δίνεται από τη σχέση (1.10).

ii. Από (1.7) και (1.8) είναι φανερό ότι, για κάθε συμμετρικό πίνακα $A \in M_n(\mathbb{F})$ η νόρμα γραμμής $|||A|||_\infty$ ταυτίζεται με τη νόρμα στήλης $|||A|||_1$.

iii. Αν ο $A \in M_n(\mathbb{R})$ είναι συμμετρικός, τότε $\rho(A) = |||A|||_2$.

iv. $|||A|||_2 = |||\,|A|\,|||_2$. □

Παράδειγμα 1.2

Έστω ο πίνακας

$$A = \begin{bmatrix} 1 & 0.098 & 0.145 \\ 0.098 & 1 & 0.019 \\ 0.145 & 0.019 & 1 \end{bmatrix},$$

στον οποίο θέλουμε να υπολογίσουμε τις νόρμες που δίνονται από τις σχέσεις (1.7) - (1.11).

- Σύμφωνα με τον τύπο (1.7), ο υπολογισμός της $|||A|||_\infty$ γίνεται ως εξής:

$$\begin{aligned} |||A|||_\infty &= \max_{1\le i\le 3}\{r_i = \sum_{j=1}^{3} |a_{ij}|\} = \max_{1\le i\le 3}\{r_i\} = \max\{r_1, r_2, r_3\} \\ &= \max\left\{|a_{11}+a_{12}+a_{13}|, |a_{21}+a_{22}+a_{23}|, |a_{31}+a_{32}+a_{33}|\right\} \\ &= \max\left\{1.243, 1.117, 1.164\right\} = 1.243 \end{aligned}$$

- Σύμφωνα με τον τύπο (1.8), ο υπολογισμός της $|||A|||_1$ γίνεται ως εξής:

$$\begin{aligned} |||A|||_1 &= \max_{1\le j\le 3}\{c_i = \sum_{i=1}^{3} |a_{ij}|\} = \max_{1\le j\le 3}\{c_j\} = \max\{c_1, c_2, c_3\} \\ &= \max\left\{|a_{11}+a_{21}+a_{31}|, |a_{12}+a_{22}+a_{32}|, |a_{13}+a_{23}+a_{33}|\right\} \\ &= \max\left\{1.243, 1.117, 1.164\right\} = 1.243 \end{aligned}$$

Αποτέλεσμα που επαληθεύει το Σχόλιο 1.1 (ii), δηλαδή $|||A|||_\infty = |||A|||_1$

i. Για τον υπολογισμό του $|||A|||_E$ χρησιμοποιώντας τον τύπο (1.9) έχουμε:

$$|||A|||_E = \sqrt{1^2 + 0.098^2 + 0.145^2 + 0.098^2 + 1^2 + 0.019^2 + 0.145^2 + 0.019^2 + 1^2} = 1.7499$$

- Σύμφωνα με το Σχόλιο 1.1 ισχύει: $|||A|||_F = |||A|||_E = 1.7499$

- Σύμφωνα με τον τύπο (1.11) και τον ορισμό της φασματικής ακτίνας στην (1.2), ο υπολογισμός της $|||A|||_2$ απαιτεί τον υπολογισμό των ιδιοτιμών του πίνακα

$$A^T A = \begin{bmatrix} 1.03 & 0.19 & 0.29 \\ 0.19 & 1.01 & 0.05 \\ 0.29 & 0.05 & 1.02 \end{bmatrix}$$

που είναι

$$\sigma(A^{\mathrm{T}} A) = \{\lambda_1 = 0.6946 \quad \lambda_2 = 0.9651 \quad \lambda_3 = 1.4023\},$$

από όπου είναι φανερό ότι $\rho(A^T A) = \max\{|\lambda_i| : \lambda_i \in \sigma(A^T A)\} = 1.4023$.

Επομένως, $|||A|||_2 = \sqrt{\rho(A^T A)} = \sqrt{1.4023} = 1.1842$ ◊

1.3 Η οργάνωση των δεδομένων

Τα δεδομένα που προκύπτουν κατά την ανάλυση μιας σειράς μετρήσεων ή παρατηρήσεων που έγιναν σε διάφορα θέματα συχνά επιβάλλεται να οργανωθούν και να εμφανιστούν με διάφορους τρόπους [10, 15, 25]. Ένας από τους τρόπους οργάνωσης των δεδομένων είναι σε μορφή πίνακα.

Τα πολυμεταβλητά δεδομένα προκύπτουν κάθε φορά που κάποιος ερευνητής επιδιώκοντας να κατανοήσει ένα κοινωνικό ή φυσικό φαινόμενο, επιλέγει να καταγράψει έναν αριθμό $p \geq 1$ μεταβλητών.

Οι τιμές των μεταβλητών καταγράφονται για κάθε διαφορετική παρατήρηση, πειραματική μονάδα ή άτομο. Με x_{kl} συμβολίζεται η τιμή της l-στής μεταβλητής η οποία παρατηρείται στη δοκιμή k. Συνεπώς, n παρατηρήσεις σε p μεταβλητές μπορούν να αναπαρασταθούν ως ακολούθως:

$$
\begin{array}{lccccccc}
 & \text{μεταβλητή}\mathbf{1} & \text{μεταβλητή}\mathbf{2} & \ldots & \text{μεταβλητή}\boldsymbol{l} & \ldots & \text{μεταβλητή}\boldsymbol{p} \\
\text{παρατήρηση}1: & x_{11} & x_{12} & \ldots & x_{1l} & \cdots & x_{1p} \\
\text{παρατήρηση}2: & x_{21} & x_{22} & \cdots & x_{2l} & \cdots & x_{2p} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
\text{παρατήρηση}\text{k}: & x_{k1} & x_{k2} & \cdots & x_{kl} & \cdots & x_{kp} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
\text{παρατήρηση}n: & x_{n1} & x_{n2} & \cdots & x_{nl} & \cdots & x_{np}
\end{array}
$$

ή να αναπαρασταθούν ως στοιχεία ενός $n \times p$ πίνακα, δηλαδή

$$
\boldsymbol{X} = \begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1l} & \cdots & x_{1p} \\
x_{21} & x_{22} & \cdots & x_{2l} & \cdots & x_{2p} \\
\vdots & \vdots & & \vdots & & \vdots \\
x_{j1} & x_{j2} & \cdots & x_{jl} & \cdots & x_{jp} \\
\vdots & \vdots & & \vdots & & \vdots \\
x_{n1} & x_{n2} & \cdots & x_{nl} & \cdots & x_{np}
\end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_l & \cdots & \boldsymbol{X}_p \end{bmatrix}
$$

1.4 Βασικές έννοιες πολυμεταβλητής στατιστικής

Ο υπολογισμός απλών περιγραφικών στατιστικών δεικτών για μονομεταβλητά δεδομένα, όπως η μέση τιμή ή η διασπορά, είναι μια διαδικασία αρκετά οικεία. Το ίδιο ενδεχομένως συμβαίνει και με τον υπολογισμό των διμεταβλητών συσχετίσεων και των αντίστοιχων συνδιακυμάνσεων που προκύπτουν από ζεύγη μετρήσεων. Τι συμβαίνει όμως κατά τον υπολογισμό περιγραφικών στατιστικών δεικτών σε ένα πολυμεταβλητό σύνολο δεδομένων;

1.4.1 Μέση τιμή

Ορισμός 1.15

Η **μέση τιμή** μιας μεταβλητής $\boldsymbol{X}_j = \begin{bmatrix} x_{1j} & x_{2j} & \cdots & x_{nj} \end{bmatrix}^T$, όπου $j = 1, 2, \ldots, p$, δίνεται από τη σχέση:

$$E(\boldsymbol{X}_j) = \overline{\mathrm{x}}_j = \frac{x_{1j} + x_{2j} + \ldots + x_{nj}}{n} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{1.12}$$

Η μέση τιμή του πίνακα X για p μεταβλητές, που βασίζεται σε np το πλήθος παρατηρήσεις, δίνεται από τη σχέση

$$\boldsymbol{\mu}^T = \begin{bmatrix} \overline{x}_1 & \overline{x}_2 & \ldots & \overline{x}_p \end{bmatrix} \tag{1.13}$$

όπου $\overline{\mathrm{x}}_j$ δίνεται από την (1.12) για κάθε $j = 1, 2, \ldots, p$.

Αρκετές ιδιότητες της μέσης τιμής διατυπώνονται σε οποιοδήποτε σύγγραμμα Στατιστικής, που αναφέρεται στη βιβλιογραφία, [8, 15, 18, 19, 25]. Οι ιδιότητες που είναι απαραίτητες στην παρούσα πτυχιακή εργασία διατυπώνονται στην ακόλουθη πρόταση.

Πρόταση 1.4

i. *Έστω ότι* $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ *είναι μεταβλητές και* $a_1, a_2, \ldots, a_p$ *σταθερές πραγματικές τιμές. Τότε:*

$$E(a_1\boldsymbol{X}_1 + a_2\boldsymbol{X}_2 + \ldots + a_p\boldsymbol{X}_p) = \sum_{j=1}^{p} a_j E(\boldsymbol{X}_j)$$

ii. *Έστω* $\boldsymbol{X}_j = \begin{bmatrix} b & b & \cdots & b \end{bmatrix}^T$ *μεταβλητή για κάθε* $j = 1, 2, \ldots, p$ και b *σταθερή πραγματική τιμή. Τότε* $E(\boldsymbol{X}_j) = b$.

iii. *Έστω* $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ *μεταβλητές, τότε:*

$$E(\boldsymbol{X}_1 + \boldsymbol{X}_2 + \ldots + \boldsymbol{X}_p) = \sum_{j=1}^{p} E(\boldsymbol{X}_j)$$

Απόδειξη:

i. Χρησιμοποιώντας πράξεις διανυσμάτων και την (1.12) μπορούμε να γράψουμε

$$\begin{aligned} E(a_1\boldsymbol{X}_1 + a_2\boldsymbol{X}_2 + \cdots + a_p\boldsymbol{X}_p) &= E(a_1\boldsymbol{X}_1) + E(a_2\boldsymbol{X}_2) + \cdots + E(a_p\boldsymbol{X}_p) \\ &= a_1 E(\boldsymbol{X}_1) + a_2 E(\boldsymbol{X}_2) + \cdots + a_p E(\boldsymbol{X}_p) \\ &= \sum_{j=1}^{p} a_j E(\boldsymbol{X}_j) \end{aligned}$$

το οποίο αποδεικνύει την ιδιότητα.

ii. Χρησιμοποιώντας την (1.12) και τη μορφή της μεταβλητής $\boldsymbol{X}_j$ μπορούμε να γράψουμε

$$E(\boldsymbol{X}_j) = E(\begin{bmatrix} b & b & \cdots & b \end{bmatrix}^T) = \frac{b + b + \ldots + b}{n} = \frac{1}{n} nb = b \ .$$

iii. Η απόδειξη είναι άμεση από την παραπάνω ιδιότητα (i) αν θεωρήσουμε $a_1 = a_2 = \cdots = a_p = 1$. □

1.4.2 Διασπορά

Ορισμός 1.16

Η **διασπορά πληθυσμού** μιας μεταβλητής $\boldsymbol{X}_j = \begin{bmatrix} x_{1j} & x_{2j} & \cdots & x_{nj} \end{bmatrix}^T$, όπου $j = 1, 2, \ldots, p$, είναι ένα μέτρο το οποίο εκφράζει τη μέση απόσταση των τιμών της από τη μέση τιμή τους και δίνεται από τη σχέση

$$\mathrm{var}(\boldsymbol{X}_j) = \sigma_j^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{\mathrm{x}}_j)^2 \tag{1.14}$$

όπου $\overline{\mathrm{x}}_j$ είναι η μέση τιμή της μεταβλητής $\boldsymbol{X}_j$ και δίνεται από την (1.12) .

Η **διασπορά δείγματος** μιας μεταβλητής $\boldsymbol{X}_j = \begin{bmatrix} x_{1j} & x_{2j} & \cdots & x_{nj} \end{bmatrix}^T$, όπου $j = 1, 2, \ldots, p$, δίνεται από τη σχέση

$$s_{jj} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \overline{\mathrm{x}}_j)^2 \tag{1.15}$$

όπου $\overline{\mathrm{x}}_j$ είναι η μέση τιμή της μεταβλητής $\boldsymbol{X}_j$ και δίνεται από την (1.12).

Η τετραγωνική ρίζα της διασποράς ονομάζεται **τυπική απόκλιση** (standard deviation). Όταν πρόκειται για πληθυσμό θα συμβολίζεται με $\mathrm{sd}_p(\boldsymbol{X}_j) = \sqrt{\mathrm{var}(\boldsymbol{X}_j)}$, όπου $\boldsymbol{X}_j$ μεταβλητή για κάθε $j = 1, 2, \ldots, p$, ενώ για δείγμα θα σημειώνεται με $sd_s(\boldsymbol{X}_j) = \sqrt{s_{jj}}$.

Σχόλιο 1.2

Οι ποσότητες $\mathrm{var}(\boldsymbol{X}_j), s_{jj} \neq 0$ με $j = 1, 2, \ldots, p$, διότι αποτελούνται από αθροίσματα τετραγώνων που τουλάχιστον ένας όρος τους είναι διάφορος του μηδενός. □

Αρκετές ιδιότητες της διασποράς διατυπώνονται σε οποιοδήποτε σύγγραμμα Στατιστικής, που αναφέρεται στη βιβλιογραφία, [8, 15, 18, 19, 25]. Οι ιδιότητες που είναι απαραίτητες στην παρούσα πτυχιακή εργασία διατυπώνονται στην ακόλουθη πρόταση.

Πρόταση 1.5

i. *Έστω* $\boldsymbol{X}_j = \begin{bmatrix} b & b & \cdots & b \end{bmatrix}^T$ *όπου* b *μια σταθερή πραγματική τιμή. Η* $\operatorname{var}(\boldsymbol{X}_j) = 0$, για $j = 1, 2, \ldots, p$.

ii. *Έστω μεταβλητή* $\boldsymbol{X}_j$ *με* $j = 1, 2, \ldots, p$. *Τότε* $\operatorname{var}(a\boldsymbol{X}_j) = a^2 \operatorname{var}(\boldsymbol{X}_j)$, *για* $a \in \mathbb{R}$.

iii. *Η διασπορά μιας μεταβλητής* $\boldsymbol{X}_j$ *ισούται με*

$$\operatorname{var}(\boldsymbol{X}_j) = \frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 - \left(E(\boldsymbol{X}_j)\right)^2 \quad \text{(πληθυσμός)}$$

$$s_{jj} = \frac{1}{n-1}\sum_{i=1}^{n} x_{ij}^2 - \frac{n}{n-1}\left(E(\boldsymbol{X}_j)\right)^2 \quad \text{(δείγμα).}$$

Απόδειξη:

i. Από τη μορφή της μεταβλητής $\boldsymbol{X}_j$ σύμφωνα με την ιδιότητα (ii) της Πρότασης 1.4 έχουμε $E(\boldsymbol{X}_j) = \overline{\mathrm{x}}_j = b$, οπότε αντικαθιστώντας στην (1.14) $x_{ij} = \overline{\mathrm{x}}_j = b$ είναι φανερό ότι $\operatorname{var}(\boldsymbol{X}_j) = 0$.

ii. Από την (1.12) και την ιδιότητα (i) της Πρότασης 1.4 ισχύει $E(a\boldsymbol{X}_j) = aE(\boldsymbol{X}_j) = a\overline{x}_j$. Αντικαθιστώντας στην (1.14) έχουμε:

$$\operatorname{var}(a\boldsymbol{X}_j) = \frac{1}{n}\sum_{i=1}^{n}(ax_{ij} - a\overline{\mathrm{x}}_j)^2 = \frac{1}{n}a^2\sum_{i=1}^{n}(x_{ij} - \overline{\mathrm{x}}_j)^2 = a^2 \operatorname{var}(\boldsymbol{X}_j)$$

iii. Χρησιμοποιώντας την (1.12), $E(\boldsymbol{X}_j) = \overline{\mathrm{x}}_j$, ο ορισμός της διασποράς του πληθυσμού στην (1.14) δίνει:

$$\operatorname{var}(\boldsymbol{X}_j) = \frac{1}{n}\sum_{i=1}^{n}(x_{ij}^2 - 2\overline{\mathrm{x}}_j x_{ij} + \overline{\mathrm{x}}_j^2) = \frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 - \frac{2\overline{\mathrm{x}}_j}{n}\sum_{i=1}^{n} x_{ij} + \frac{n\overline{\mathrm{x}}_j^2}{n} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 - 2\overline{\mathrm{x}}_j^2 + \overline{\mathrm{x}}_j^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 - \left(E(\boldsymbol{X}_j)\right)^2$$

Όμοια αποδεικνύεται και η επόμενη ισότητα για τη διασπορά του δείγματος χρησιμοποιώντας την (1.15). □

1.4.3 Συνδιακύμανση

Ορισμός 1.17

Η **συνδιακύμανση πληθυσμού** (population covariance) δυο μεταβλητών $\boldsymbol{X}_i = \begin{bmatrix} x_{1i} & x_{2i} & \cdots & x_{ni} \end{bmatrix}^T$ και $\boldsymbol{X}_j = \begin{bmatrix} x_{1j} & x_{2j} & \cdots & x_{nj} \end{bmatrix}^T$, όπου $i, j = 1, 2, \ldots, p$, δίνεται από τη σχέση:

$$C\text{ov}_p(\boldsymbol{X}_i, \boldsymbol{X}_j) = \frac{1}{n}\sum_{k=1}^{n}(x_{ki} - \overline{\text{x}}_i)(x_{kj} - \overline{\text{x}}_j) \qquad (1.16)$$

όπου $\overline{\text{x}}_i$ και $\overline{\text{x}}_j$ είναι οι μέσες τιμές των μεταβλητών $\boldsymbol{X}_i$ και $\boldsymbol{X}_j$, αντίστοιχα.

Η **συνδιακύμανση δείγματος** (sample covariance) δυο μεταβλητών $\boldsymbol{X}_i = \begin{bmatrix} x_{1i} & x_{2i} & \cdots & x_{ni} \end{bmatrix}^T$ και $\boldsymbol{X}_j = \begin{bmatrix} x_{1j} & x_{2j} & \cdots & x_{nj} \end{bmatrix}^T$, όπου $i, j = 1, 2, \ldots, p$, δίνεται από τη σχέση:

$$s_{ij} = \text{cov}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \frac{1}{n-1}\sum_{k=1}^{n}(x_{ki} - \overline{\text{x}}_i)(x_{kj} - \overline{\text{x}}_j) \qquad (1.17)$$

όπου $\overline{\text{x}}_i$ και $\overline{\text{x}}_j$ είναι οι μέσες τιμές των μεταβλητών $\boldsymbol{X}_i$ και $\boldsymbol{X}_j$, αντίστοιχα.

Ειδική περίπτωση της συνδιακύμανσης αποτελεί η διασπορά, επειδή για $i = j$ η ισότητα στην (1.16) ταυτίζεται με την (1.14) και η (1.17) με την (1.15), όπως παρουσιάζεται και στην ιδιότητα (i) της Πρότασης 1.6 ακολούθως.

Στην (1.16) και (1.17) η αντιμεταθετικότητα που ισχύει στους πραγματικούς αριθμούς μας επιτρέπει να συμπεράνουμε $C\text{ov}_p(\boldsymbol{X}_i, \boldsymbol{X}_j) = C\text{ov}_p(\boldsymbol{X}_j, \boldsymbol{X}_i)$ και $s_{ij} = s_{ji}$ για κάθε $i, j = 1, 2, \ldots, p$.

Οι σημαντικότερες ιδιότητες της συνδιακύμανσης διατυπώνονται στην ακόλουθη πρόταση, η απόδειξη των οποίων μπορεί να αναζητηθεί σε οποιοδήποτε σύγγραμμα Στατιστικής, που αναφέρεται στη βιβλιογραφία, [8, 15, 18, 19, 25].

Πρόταση 1.6

i. $\mathrm{Cov}_p(\boldsymbol{X}_i, \boldsymbol{X}_i) = \mathrm{var}(\boldsymbol{X}_i)$ *και* $\mathrm{cov}(\boldsymbol{X}_i, \boldsymbol{X}_i) = s_{ii}$, *για κάθε* $i = 1, 2, \ldots, p$.

ii. *Έστω* $\boldsymbol{X}_i, \boldsymbol{X}_j$ *μεταβλητές και* a_1, a_2 *πραγματικές σταθερές τιμές. Τότε:*

$$\mathrm{Cov}_p(a_1\boldsymbol{X}_i, a_2\boldsymbol{X}_j) = a_1 a_2 \mathrm{Cov}_p(\boldsymbol{X}_i, \boldsymbol{X}_j)$$

$$\mathrm{cov}(a_1\boldsymbol{X}_i, a_2\boldsymbol{X}_j) = a_1 a_2\, \mathrm{cov}(\boldsymbol{X}_i, \boldsymbol{X}_j)$$

iii. $\mathrm{var}(a_1\boldsymbol{X}_i + a_2\boldsymbol{X}_j) = a_1^{\ 2}\,\mathrm{var}(\boldsymbol{X}_i) + a_2^{\ 2}\,\mathrm{var}(\boldsymbol{X}_j) + 2a_1a_2\mathrm{Cov}_p(\boldsymbol{X}_i, \boldsymbol{X}_j)$

iv. *Εάν η μεταβλητή* $\boldsymbol{X}_i$ *είναι ανεξάρτητη της μεταβλητής* $\boldsymbol{X}_j$, *τότε*

$$\mathrm{Cov}_p(\boldsymbol{X}_i, \boldsymbol{X}_j) = 0,\ \mathrm{cov}(\boldsymbol{X}_i, \boldsymbol{X}_j) = 0 \ \textit{και}\ \mathrm{var}(\boldsymbol{X}_i + \boldsymbol{X}_j) = \mathrm{var}(\boldsymbol{X}_i) + \mathrm{var}(\boldsymbol{X}_j)$$

Οι συνδιακυμάνσεις όλων των δυνατών συνδυασμών μεταξύ των μεταβλητών μπορούν να τοποθετηθούν σε έναν πίνακα, όπως διατυπώνεται στον ακόλουθο ορισμό.

Ορισμός 1.18

Ο πίνακας S στην (1.18) ονομάζεται **πίνακας συνδιακύμανσης** των μεταβλητών $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$,

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{12} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_{pp} \end{bmatrix} \tag{1.18}$$

όπου τα στοιχεία s_{ij} , $i, j = 1, 2, \ldots, p$ υπολογίζονται από (1.15) και (1.17) .

Προφανώς σύμφωνα με τους Ορισμούς 1.16, 1.17 και 1.18 ο πίνακας συνδιακύμανσης έχει πραγματικά στοιχεία και είναι συμμετρικός. Στη συνέχεια, παρουσιάζονται οι ιδιότητες του πίνακα συνδιακύμανσης S οι οποίες μπορούν εύκολα να αποδειχθούν χρησιμοποιώντας τις ιδιότητες (vii) και (viii) της Πρότασης 1.3 και τα Θεωρήματα 1.1 και 1.2.

Πρόταση 1.7

i. *Ο πίνακας συνδιακύμανσης* S *είναι συμμετρικός με πραγματικές*[11] *ιδιοτιμές και θετικά ημιορισμένος*[12]. *Η διάταξη των ιδιοτιμών είναι η ακόλουθη*

$$0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \ldots \leq \lambda_p \,.$$

ii. *Ο πίνακας* S *διαγωνοποιείται*[13] *από ορθογώνιο πίνακα* U *με διαγώνια μορφή:*

$$S = U\Delta_S U^T \tag{1.19}$$

όπου $\Delta_S = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$.

---

[11] Βλέπε ιδιότητα (vii), Πρόταση 1.3.

[12] Βλέπε ιδιότητα (viii), Πρόταση 1.3.

[13] Βλέπε Θεώρημα 1.1.

1.4.4 Συντελεστής γραμμικής συσχέτισης

Στο πλαίσιο της παρούσας πτυχιακής εργασίας θα ασχοληθούμε μόνο με τη γραμμική συσχέτιση, η οποία συνιστά την απλούστερη μορφή συσχέτισης. Ο συντελεστής γραμμικής συσχέτισης αποτελεί το μέτρο συσχέτισης μεταξύ δυο μεταβλητών $\boldsymbol{X}_i$ και $\boldsymbol{X}_j$, για $i, j = 1, 2, \ldots, p$, το οποίο δεν εξαρτάται από τις μονάδες της μέτρησης [15, 25] και εκφράζει τη συγκέντρωση των σημείων ενός διαγράμματος διασποράς γύρω από την ευθεία παλινδρόμησης μεταξύ των μεταβλητών.

Ορισμός 1.19

Ο **συντελεστής γραμμικής συσχέτισης** (correlation) των μεταβλητών $\boldsymbol{X}_i$ και $\boldsymbol{X}_j$, για $i, j = 1, 2, \ldots, p$, ορίζεται:

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}} = \frac{\sum_{k=1}^{n}\left(x_{ki} - \overline{\mathrm{x}}_i\right)\left(x_{kj} - \overline{\mathrm{x}}_j\right)}{\sqrt{\sum_{k=1}^{n}\left(x_{ki} - \overline{\mathrm{x}}_i\right)^2}\sqrt{\sum_{k=1}^{n}\left(x_{kj} - \overline{\mathrm{x}}_j\right)^2}} \tag{1.20}$$

Άμεση συνέπεια του παραπάνω ορισμού, της (1.20) και των ιδιοτήτων (i) και (iv) της Πρότασης 1.6 είναι οι ιδιότητες της συσχέτισης που παρατίθενται στην επόμενη πρόταση, η απόδειξη των οποίων μπορεί να αναζητηθεί σε οποιοδήποτε σύγγραμμα Στατιστικής [8, 18, 19].

Πρόταση 1. 8

i. $r_{ij} = r_{ji}$.

ii. *Η τιμή του συντελεστή συσχέτισης είναι μεταξύ του* -1 *και* $+1$

$$-1 \le r_{ij} \le +1$$

όπου $i, j = 1, 2, \ldots, p$.

iii. *Αν* $\boldsymbol{X}_i = \boldsymbol{X}_j$, *τότε* $r_{ij} = 1$.

iv. *Εάν* $\mathrm{s}_{ii} = \mathrm{s}_{jj} = 1$, *τότε* $r_{ij} = s_{ij}$.

v. *Εάν* $0 < r_{ij} < 1$, *τότε οι* $\boldsymbol{X}_i, \boldsymbol{X}_j$ *είναι γραμμικά θετικά συσχετισμένες.*

vi. *Εάν* $-1 < r_{ij} < 0$, *τότε οι* $\boldsymbol{X}_i, \boldsymbol{X}_j$ *είναι γραμμικά αρνητικά συσχετισμένες.*

vii. *Εάν* $r_{ij} = 0$, *τότε οι μεταβλητές* $\boldsymbol{X}_i, \boldsymbol{X}_j$ *είναι γραμμικά ασυσχέτιστες.*

viii. *Εάν* $r_{ij} = 1$, *τότε υπάρχει εξάρτηση μεταξύ των μεταβλητών* $\boldsymbol{X}_i, \boldsymbol{X}_j$ *και μάλιστα γραμμικά θετική, που δίνεται* $\boldsymbol{X}_j = a + b\boldsymbol{X}_i$, *με* $b > 0$. *Αντίστοιχα, εάν* $r_{ij} = -1$ *υπάρχει γραμμικά αρνητική εξάρτηση μεταξύ των μεταβλητών* $\boldsymbol{X}_i, \boldsymbol{X}_j$, *που δίνεται* $\boldsymbol{X}_j = a + b\boldsymbol{X}_i$ *, με* $b < 0$.

Οι συσχετίσεις όλων των δυνατών συνδυασμών μεταξύ των μεταβλητών μπορούν να τοποθετηθούν σε έναν πίνακα όπως διατυπώνεται στον ακόλουθο ορισμό.

Ορισμός 1.20

Ο πίνακας R στην (1.21) ονομάζεται **πίνακας συσχέτισης** των μεταβλητών $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$,

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix} \tag{1.21}$$

όπου τα στοιχεία r_{ij}, $i, j = 1, 2, \ldots, p$ υπολογίζονται από τον (1.20).

Προφανώς από τις ιδιότητες (i) - (iii) της Πρότασης 1.8 και τον Ορισμό 1.20 ένας πίνακας συσχέτισης έχει πραγματικά στοιχεία και είναι συμμετρικός. Στη συνέχεια, παρουσιάζονται οι ιδιότητες του πίνακα συσχέτισης R, οι οποίες μπορούν εύκολα να αποδειχθούν χρησιμοποιώντας την ιδιότητα (vii) της Πρότασης 1.3 και τα Θεωρήματα 1.1 και 1.2.

Πρόταση 1. 9

i. *Ο πίνακας συσχέτισης* R *είναι συμμετρικός*[14] *με πραγματικές*[15] *ιδιοτιμές.*

ii. *Ο πίνακας* R *διαγωνοποιείται από ορθογώνιο πίνακα από ορθογώνιο πίνακα* W *με διαγώνια μορφή:*

$$R = W\Delta_R W^T \tag{1.22}$$

όπου $\Delta_R = diag(r_1, r_2, \ldots, r_p)$.

---

[14] Βλέπε Ορισμό 1.4.

[15] Βλέπε ιδιότητα (vii), Πρόταση 1.3.

Στη συνέχεια, παρουσιάζεται η σχέση που συνδέει τον πίνακα συνδιακύμανσης S και τον πίνακα συσχέτισης R, όπως παρουσιάζεται [15].

Πρόταση 1.10

Ο πίνακας συσχέτισης R συνδέεται με τον πίνακα συνδιακύμανσης S με την ακόλουθη σχέση

$$S = V^{1/2} R V^{1/2} \tag{1.23}$$

όπου $V^{1/2}$ είναι ο $p \times p$ διαγώνιος πίνακας:

$$V^{1/2} = \begin{bmatrix} \sqrt{s_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{s_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{s_{pp}} \end{bmatrix}. \tag{1.24}$$

Προφανώς από την (1.24) και τις ιδιότητες (ii) και (vi) της Πρότασης 1.1 προκύπτουν οι ιδιότητες του $V^{1/2}$, οι οποίες παρουσιάζονται στην επόμενη πρόταση.

Πρόταση 1.11

i. *Η ορίζουσα του πίνακα $V^{1/2}$ ισούται με το γινόμενο των διαγώνιων στοιχείων του και είναι διάφορη του μηδενός*[16] *.*

ii. *Ο πίνακας $V^{1/2}$ είναι αντιστρέψιμος*[17]*.*

Άμεση συνέπεια της (1.23) της Πρότασης 1.10 και της ιδιότητας (ii) της Πρότασης 1.11 είναι το ακόλουθο πόρισμα.

Πόρισμα 1.1

Ο πίνακας συσχέτισης συνδέεται με τον πίνακα συνδιακύμανσης:

$$R = (V^{1/2})^{-1} S (V^{1/2})^{-1}$$

---

[16] Βλέπε Σχόλιο 1.2.

[17] Βλέπε ιδιότητα (ii), Πρόταση 1.1

Αξιοποιώντας τα αποτελέσματα του Πορίσματος 1.1 ο πίνακας συσχέτισης R θα υπολογίζεται άμεσα από τον πίνακα S και για αυτό στην παρούσα πτυχιακή το βάρος πέφτει στον υπολογισμό του S .

Πρόταση 1.12

Οι ιδιοτιμές του πίνακα συσχέτισης R συνδέονται με τις ιδιοτιμές του αντίστοιχου πίνακα συνδιακύμανσης S :

$$\Delta_R = W^T (V^{1/2})^{-1} U \Delta_S (V^{1/2} U)^{-1} W \tag{1.25}$$

Απόδειξη:

Αντικαθιστώντας τον S από την (1.19) της Πρότασης 1.7 στη σχέση του Πορίσματος 1.1 και λαμβάνοντας υπόψη ότι $U^{-1} = U^T$, έχουμε:

$$R = (V^{1/2})^{-1} U \Delta_S U^T (V^{1/2})^{-1} = (V^{1/2})^{-1} U \Delta_S U^{-1} (V^{1/2})^{-1} = (V^{1/2})^{-1} U \Delta_S (V^{1/2} U)^{-1}$$

Στη συνέχεια, αντικαθιστώντας στην τελευταία ισότητα τον πίνακα συσχέτισης R από την (1.22) και λαμβάνοντας υπόψη ότι $W^{-1} = W^T$, έχουμε:

$$W \Delta_R W^T = (V^{1/2})^{-1} U \Delta_S (V^{1/2} U)^{-1} \Rightarrow \Delta_R = W^T (V^{1/2})^{-1} U \Delta_S (V^{1/2} U)^{-1} W$$

□

Η ακόλουθη πρόταση δίνει πληροφορίες για το πρόσημο των ιδιοτιμών του πίνακα συσχέτισης.

Πρόταση 1.13

Ο πίνακας συσχέτισης R είναι θετικά ημιορισμένος πίνακας με ιδιοτιμές που έχουν τη διάταξη:

$$0 \le r_1 \le r_2 \le \ldots \le \ldots \le r_p \, .$$

Απόδειξη:

Επειδή ο πίνακας συνδιακύμανσης S έχει μη αρνητικές ιδιοτιμές, που έχουν την διάταξη όπως εμφανίζεται στην ιδιότητα (i) της Πρόταση 1.7, χρησιμοποιώντας και τη σχέση (1.25) της Πρότασης 1.12 συμπεραίνουμε ότι ισχύει το ίδιο και για τον πίνακα συσχέτισης R .

□

Παράδειγμα 1.3

Έστω τρεις μεταβλητές $\boldsymbol{X}_1, \boldsymbol{X}_2$ και $\boldsymbol{X}_3$ οι οποίες έχουν τιμές όπως αναγράφονται στον ακόλουθο πίνακα:

$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$
1	2	14
3	1	2
4	5	1
8	2	2
0	3	3
5	7	8
2	8	1
3	0	3
4	1	4
0	1	2

Να βρεθεί η μέση τιμή, ο πίνακας συνδιακύμανσης και συσχέτισης για το παραπάνω δείγμα και να επαληθευθεί το Πόρισμα 1.1.

Αρχικά, υπολογίζουμε τη μέσης τιμή κάθε μεταβλητής $\boldsymbol{X}_1, \boldsymbol{X}_2$ και $\boldsymbol{X}_3$ με τη χρήση της σχέσης (1.12) :

$$\overline{\mathrm{x}}_1 = E(X_1\ ) = \frac{1+3+4+8+0+5+2+3+4+0}{10} = 3$$

$$\overline{\mathrm{x}}_2 = E(X_2\ ) = \frac{2+1+5+2+3+7+8+0+1+1}{10} = 3$$

$$\overline{\mathrm{x}}_3 = E(X_3)\ = \frac{14+2+1+2+3+8+1+3+4+2}{10} = 4$$

Στη συνέχεια, υπολογίζουμε τα στοιχεία του πίνακα συνδιακύμανσης βάσει της (1.17) :

$$s_{11} = \frac{1}{10-1}\sum_{i=1}^{10}(x_1 - \overline{\mathrm{x}}_1)^2$$

$$= \frac{(1-3)^2+(3-3)^2+(4-3)^2+\ldots+(3-3)^2+(4-3)^2+(0-3)^2}{9} = \frac{54}{9} = 6$$

$$s_{22} = \frac{1}{10-1}\sum_{i=1}^{10}(x_2 - \overline{\mathrm{x}}_2)^2 = \frac{68}{9} = 7.555$$

$$s_{33} = \frac{1}{10-1}\sum_{i=1}^{10}(x_3 - \bar{x}_3)^2 = \frac{148}{9} = 16.444$$

$$s_{12} = \frac{1}{10-1}\sum_{j=1}^{10}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) = \frac{6}{9} = 0.666$$

$$s_{13} = \frac{1}{10-1}\sum_{j=1}^{10}(x_1 - \bar{x}_1)(x_3 - \bar{x}_3) = \frac{-13}{9} = -1.444$$

$$s_{23} = \frac{1}{10-1}\sum_{j=1}^{10}(x_2 - \bar{x}_2)(x_3 - \bar{x}_3) = \frac{-2}{2} = -0.222$$

Συνεπώς, από την (1.13) η μέση τιμή και από την (1.18) ο πίνακας συνδιακύμανσης είναι:

$$\bar{\boldsymbol{\mu}}^T = \begin{bmatrix} E(\boldsymbol{X}_1) & E(\boldsymbol{X}_2) & E(\boldsymbol{X}_3) \end{bmatrix} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \bar{x}_3 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 4 \end{bmatrix}$$

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix} = \begin{bmatrix} 6 & 0.666 & -1.444 \\ 0.666 & 7.555 & -0.222 \\ -1.444 & -0.222 & 16.444 \end{bmatrix}$$

Ο πίνακας συνδιακύμανσης S που προέκυψε είναι συμμετρικός επαληθεύοντας την ιδιότητα (i) της Πρότασης 1.7.

Οι ιδιοτιμές του πίνακα συνδιακύμανσης υπολογίζονται από τις ρίζες της χαρακτηριστικής εξίσωσης στην (1.3), που είναι Αναλυτικά:

$$\det(S - \lambda I) = 0 \Rightarrow -\lambda^3 + 29.999\lambda^2 - 265.6504 + 722.491 = 0$$

Άρα οι ιδιοτιμές του πίνακα συνδιακύμανσης S είναι

$$\lambda_1 = 5.6,\ \lambda_2 = 7.748,\ \lambda_3 = 16.651,$$

γεγονός που επαληθεύει την ιδιότητα (i) της Πρότασης 1.7.

Χρησιμοποιώντας την (1.20) και τις παραπάνω διασπορές/συνδιακυμάνσεις υπολογίζονται όλες οι συσχετίσεις των τριών μεταβλητών ως ακολούθως:

$$r_{11} = \frac{s_{11}}{\sqrt{s_{11}}\sqrt{s_{11}}} = \frac{6}{\sqrt{6}\sqrt{6}} = 1$$

$$r_{22} = \frac{s_{22}}{\sqrt{s_{22}}\sqrt{s_{22}}} = \frac{7.555}{\sqrt{7.555}\sqrt{7.555}} = 1$$

$$r_{33} = \frac{s_{33}}{\sqrt{s_{33}}\sqrt{s_{33}}} = \frac{16.444}{\sqrt{16.444}\sqrt{16.444}} = 1$$

$$r_{12} = r_{21} = \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} = \frac{0.666}{\sqrt{6}\sqrt{7.555}} = 0.099$$

$$r_{13} = r_{31} = \frac{s_{13}}{\sqrt{s_{11}}\sqrt{s_{33}}} = \frac{-1.444}{\sqrt{6}\sqrt{16.444}} = -0.145$$

$$r_{23} = r_{32} = \frac{s_{23}}{\sqrt{s_{22}}\sqrt{s_{33}}} = \frac{0.222}{\sqrt{7.555}\sqrt{16.444}} = -0.019$$

Συνεπώς, ο πίνακας συσχέτισης είναι:

$$R = \begin{bmatrix} 1 & 0.099 & -0.145 \\ 0.09 & 1 & -0.019 \\ -0.145 & -0.019 & 1 \end{bmatrix}$$

Για την επαλήθευση του Πορίσματος 1.1 έχουμε:

$$(V^{1/2})^{-1} S (V^{1/2})^{-1} =$$

$$= \begin{bmatrix} \frac{1}{\sqrt{6}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{7.555}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{16.444}} \end{bmatrix} \begin{bmatrix} 6 & 0.666 & -1.444 \\ 0.666 & 7.555 & -0.222 \\ -1.444 & -0.222 & 16.444 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{7.555}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{16.444}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.099 & -0.145 \\ 0.09 & 1 & -0.019 \\ -0.145 & -0.019 & 1 \end{bmatrix} = R$$

◊

1.5 Μοντέλο γραμμικής παλινδρόμησης

Η ανάλυση παλινδρόμησης είναι μια στατιστική μεθοδολογία για την πρόβλεψη τιμών μιας ή περισσοτέρων εξαρτημένων μεταβλητών από ένα σύνολο ανεξάρτητων μεταβλητών. Επίσης, μπορεί να χρησιμοποιηθεί για την αξιολόγηση των αποτελεσμάτων των ανεξάρτητων μεταβλητών από τις εξαρτημένες. Δυστυχώς, η παλινδρόμηση πήρε το όνομα της από τον τίτλο της πρώτης ερευνητικής εργασίας της σχετικής με αυτό το θέμα [11], το οποίο δεν αντιπροσωπεύει ούτε τη σημαντικότητα αλλά ούτε το εύρος της εφαρμογής της.

1.5.1 Γενικό πλαίσιο γραμμικής παλινδρόμησης

Το κλασσικό μοντέλο γραμμικής παλινδρόμησης ασχολείται με τη σύνδεση μεταξύ μιας εξαρτημένης μεταβλητής $\boldsymbol{Y}$ και μιας συλλογής από προβλεπτικές μεταβλητές $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$. Το μοντέλο της παλινδρόμησης μεταχειρίζεται την $\boldsymbol{Y}$ ως μια μεταβλητή της οποίας η μέση τιμή εξαρτάται από σταθερές τιμές των x_i. Αυτή η μέση τιμή υποτίθεται ότι είναι μια γραμμική συνάρτηση των συντελεστών παλινδρόμησης (regression coefficients) $b_0, b_1, \ldots, b_p$.

Το μοντέλο γραμμικής παλινδρόμησης επίσης προκύπτει σαν μια διαφορά δεδομένων. Υποθέτουμε ότι όλες οι μεταβλητές $\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ έχουν κοινή κατανομή, όχι απαραίτητα κανονική, με $(p+1)\times 1$ διάνυσμα μέσης τιμής $\boldsymbol{\mu}$ και έναν $(p+1)\times(p+1)$ πίνακα συνδιακύμανσης S. Διαμερίζοντας τους πίνακες $\boldsymbol{\mu}$ και S με έναν προφανή τρόπο, γράφουμε:

$$\boldsymbol{\mu} = \begin{bmatrix} \bar{\mathrm{Y}} \\ \bar{\boldsymbol{X}} \end{bmatrix} \text{ και } S = \begin{bmatrix} s_{11} & S_{12} \\ S_{12}^T & S_{22} \end{bmatrix}$$

όπου

$\bar{\mathrm{Y}} = E(\boldsymbol{Y})$ είναι ένας 1×1 πίνακας,

$\bar{\mathrm{X}} = E(\boldsymbol{X})$ είναι ένας $p\times 1$ πίνακας,

$s_{11} = s_{YY}$,

$\boldsymbol{S}_{12} = \left[\mathrm{cov}(\boldsymbol{X}_1,\boldsymbol{Y}) \quad \mathrm{cov}(\boldsymbol{X}_2,\boldsymbol{Y}) \quad \cdots \quad \mathrm{cov}(\boldsymbol{X}_p,\boldsymbol{Y})\right]$ είναι ένας $1\times p$ πίνακας ,

S_{22} είναι ο $p \times p$ πίνακας συνδιακύμανσης των μεταβλητών $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ της (1.18).

Εξετάζοντας το πρόβλημα της πρόβλεψης της μεταβλητής $\boldsymbol{Y}$ χρησιμοποιώντας τη σχέση:

$$\hat{\mathrm{Y}} = b_0 + b_1\boldsymbol{X}_1 + \ldots + b_p\boldsymbol{X}_p = b_0 + b'X \qquad (1.26)$$

Πρόταση 1.14

Έστω $\boldsymbol{\mu} = \begin{bmatrix} \overline{\mathrm{Y}} \\ \overline{\boldsymbol{X}} \end{bmatrix}$ *και* $S = \begin{bmatrix} s_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{12}^T & S_{22} \end{bmatrix}$ *να είναι το διάνυσμα μέσης τιμής και ο πίνακας συνδιακύμανσης, αντίστοιχα, για ένα τυχαίο δείγμα μεγέθους* n. *Τότε οι εκτιμητές της μέγιστης πιθανοφάνειας των συντελεστών της γραμμικής πρόβλεψης είναι:*

$$b = S_{22}^{-1}\boldsymbol{S}_{12}^T \qquad (1.27)$$

$$b_0 = \overline{\mathrm{Y}} - \boldsymbol{S}_{12}S_{22}^{-1}\overline{X} = \overline{\mathrm{Y}} - b^{'}\overline{X} \qquad (1.28)$$

Παράδειγμα 1.4

Υπολογισμός των αγνώστων παραμέτρων $\boldsymbol{b}$ της σχέσης (1.26) για τα δεδομένα του Παραδείγματος 1.3, το $n = 10$.

Το διάνυσμα μέσης τιμής ισούται με: $\boldsymbol{\mu} = \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}$

Ο πίνακας συνδιακύμανσης είναι ο: $S = \begin{bmatrix} 6 & 0.66 & 1.44 \\ 0.66 & 7.5 & 0.22 \\ 1.44 & 0.22 & 16.4 \end{bmatrix}$

Με τη χρήση των σχέσεων (1.27), (1.28) υπολογίζουμε τους συντελεστές b και b_0.

$$\boldsymbol{b} = S_{22}^{-1}\boldsymbol{S}_{12}^T = \begin{bmatrix} 7.5 & 0.22 \\ 0.22 & 16.4 \end{bmatrix}^{-1} \begin{bmatrix} 0.66 \\ 1.44 \end{bmatrix} = \begin{bmatrix} 0.086 \\ 0.087 \end{bmatrix}$$

$$\boldsymbol{b}_0 = \overline{\mathrm{Y}} - \boldsymbol{S}_{12}S_{22}^{-1}\overline{X} = \overline{\mathrm{Y}} - \boldsymbol{b}^{'}\overline{\boldsymbol{X}} = 3 - \left(\begin{bmatrix} 0.086 & 0.087 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \right) = 2.397$$

και η συνάρτηση εκτιμώμενης παλινδρόμησης:

$$b_0 + b = 2.397 - 0.086\boldsymbol{X}_1 - 0.087\boldsymbol{X}_2 \qquad \Diamond$$

2^o ΚΕΦΑΛΑΙΟ

ΠΙΝΑΚΕΣ ΣΥΝΔΙΑΚΥΜΑΝΣΗΣ ΑΠΟ ΕΛΛΙΠΗ ΔΕΔΟΜΕΝΑ

Στο προηγούμενο κεφάλαιο κατέστει σαφής ο τρόπος υπολογισμού των πινάκων συνδιακύμανσης και συσχέτισης, οι οποίοι προέρχονται από ένα σύνολο δεδομένων όπου όλες οι παρατηρήσεις είναι πλήρεις. Τι συμβαίνει όμως στην περίπτωση που το σύνολο δεδομένων εκτός από ορισμένες πλήρεις παρατηρήσεις περιέχει και ελλιπή δεδομένα (missing data);

Τις τελευταίες δεκαετίες πολλοί ερευνητές από διάφορους επιστημονικούς κλάδους έχουν ασχοληθεί με τη μελέτη στατιστικών δεικτών, οι οποίοι προέρχονται από σύνολα δεδομένων, που περιγράφονται στο παραπάνω ερώτημα. Για παράδειγμα, ο υπολογισμός της μέσης τιμής, της διασποράς, της συνδιακύμανσης και της συσχέτισης δεν είναι εφικτό να βασιστεί στους ορισμούς που δόθηκαν στο προηγούμενο κεφάλαιο, αλλά σε μεθόδους οι οποίες θα παρατεθούν εκτενώς στη συνέχεια και οι οποίες είναι:

1. μέθοδος παρατηρήσεων με ελλιπή δεδομένα (listwise deletion) [13, 20, 22, 23]
2. μέθοδος EM (Expectation–maximization algorithm) [2, 5, 9, 15]
3. μέθοδος διαγραφής ανά ζεύγη (pairwise deletion) [3, 13, 27]

Αξίζει να σημειωθεί ότι το σύνολο δεδομένων που χρησιμοποιήθηκε για την στατιστική ανάλυση παρατίθεται στο Παράρτημα Γ. Ο κώδικας που αναπτύχθηκε για Stata προκειμένου να υπολογιστεί ο πίνακας της μεθόδου διαγραφής ανά ζεύγη παρατίθεται στο Παράρτημα Α.

2.1. Μέθοδος παρατηρήσεων με ελλιπή δεδομένα

Έστω ένα σύνολο δεδομένων που έχει καταγραφεί για p μεταβλητές έχει ορισμένες παρατηρήσεις με ελλιπή δεδομένα x_{ij}, για κάποια $i=1,...,n$, $j=1,...,p$. Η ιδέα που προτείνεται σε αρκετές ερευνητικές εργασίες [13, 20, 22, 23] είναι η διαγραφή όλων των παρατηρήσεων που έχουν ελλιπή δεδομένα, δηλαδή στην παρούσα μέθοδο μια ολόκληρη παρατήρηση εξαιρείται από την ανάλυση, αν υπάρχει τουλάχιστον ένα ελλιπές δεδομένο x_{ij} σε αυτήν. Με αυτόν τον τρόπο παράγεται ένα μειωμένου μεγέθους σύνολο δεδομένων, αφού ελαττώνεται το πλήθος των παρατηρήσεων και παραμένει σταθερό το πλήθος των μεταβλητών. Είναι προφανές ότι μετά την εξαίρεση όλων των παρατηρήσεων με ελλιπή δεδομένα, το σύνολο έχει όλες τις παρατηρήσεις πλήρεις, οπότε η εύρεση του πίνακα συνδιακύμανσης μπορεί να βασιστεί στον Ορισμό 1.18.

Στη συνέχεια της παρούσας πτυχιακής ο πίνακας συνδιακύμανσης από τη μέθοδο παρατηρήσεων με ελλιπή δεδομένα συμβολίζεται S_{LW}. Επίσης, η μέθοδος αυτή χαρακτηρίζεται ως *πλήρης ανάλυση κατά περίπτωση* (Complete Case Analysis) [13].

Σχόλια 2.1

i. Τα πλεονεκτήματα της μεθόδου παρατηρήσεων με ελλιπή δεδομένα είναι:
    a. η μείωση του συνόλου δεδομένων καθώς προσφέρει ευχρηστία, δεδομένου ότι τα στατιστικά πακέτα μπορούν πλέον εύκολα να εφαρμοστούν και συγκρισιμότητα [20].
    b. η μέθοδος ενδείκνυται σε περιπτώσεις όπου το πλήθος των παρατηρήσεων με ελλιπή δεδομένα είναι σχετικά μικρό [22].

    Από την άλλη πλευρά το βασικό μειονέκτημα της μεθόδου είναι η χαμηλότερη ακρίβεια των εκτιμήσεων που θα προκύψουν από το σύνολο δεδομένων με τις μειωμένες παρατηρήσεις λόγω του μικρότερου μεγέθους του δείγματος [13].

ii. Επειδή κατά την εφαρμογή της μεθόδου παρατηρήσεων με ελλιπή δεδομένα υπολογίστηκε ο πίνακας συνδιακύμανσης S_{LW}, σύμφωνα με το Πόρισμα 1.1 μπορεί να υπολογιστεί άμεσα και ο αντίστοιχος πίνακας συσχέτισης που θα συμβολίζεται R_{LW}.

iii. Σύμφωνα με τον υπολογισμό του S_{LW}, ο οποίος προέρχεται από σύνολο παρατηρήσεων με πλήρη δεδομένα, και αξιοποιώντας τις ιδιότητες του πίνακα συνδιακύμανσης όπως αυτές αναφέρονται στο (i) της Πρότασης 1.7, οι ιδιοτιμές του S_{LW} είναι μη αρνητικές. Συνεπώς, S_{LW} είναι **πάντα** ένας **θετικά ημιορισμένος** πίνακας.

    Ομοίως, ο αντίστοιχος R_{LW} είναι **πάντα** ένας θετικά **ημιορισμένος πίνακας** (βλέπε Πρόταση 1.13). □

Χρησιμοποιώντας ως σύνολο με πλήρη δεδομένα αυτό του Παραδείγματος 1.3 στο ακόλουθο παράδειγμα θα εφαρμόσουμε τη μέθοδο των παρατηρήσεων με ελλιπή δεδομένα μιας και το σύνολο των παρατηρήσεων είναι αρκετά μικρό (βλέπε Σχόλιο 2.1 (ib)).

Παράδειγμα 2.1

Έστω τρεις μεταβλητές $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$, οι οποίες έχουν τιμές όπως αυτές αναγράφονται στον ακόλουθο πίνακα με ελλιπή δεδομένα:

$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$
1	2	14
3	_	2
_	_	1
8	2	2
0	3	_
5	_	_
_	8	1
3	0	3
4	1	4
0	1	2

Να βρεθούν:

i. ο πίνακας συνδιακύμανσης S_{LW},

ii. ο πίνακας συσχέτισης R_{LW}, και

iii. να υπολογιστούν οι ιδιοτιμές του S_{LW}.

Σύμφωνα με τη μέθοδο των παρατηρήσεων με ελλιπή δεδομένα που αναπτύχθηκε παραπάνω θα ληφθούν υπόψη μόνο οι παρατηρήσεις που είναι πλήρεις (στον παραπάνω πίνακα σημειώνονται με γαλάζιο χρώμα). Έτσι, δημιουργείται ένα μικρότερο σύνολο δεδομένων το ακόλουθο.

$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$
1	2	14
8	2	2
3	0	3
4	1	4
0	1	2

Το παραπάνω σύνολο δεδομένων έχει την ίδια μορφή με αυτή του Παραδείγματος 1.3.

i. Ο πίνακας συνδιακύμανσης S_{LW} υπολογίζεται από την (1.18) του Ορισμού 1.18.

Οι μέσες τιμές κάθε μεταβλητής $\boldsymbol{X}_1, \boldsymbol{X}_2$ και $\boldsymbol{X}_3$ είναι:

$$\overline{\mathrm{x}}_1 = \mathrm{E}(\boldsymbol{X}_1\ ) = \frac{1+8+3+4+0}{5} = 3.2$$

$$\overline{\mathrm{x}}_2 = \mathrm{E}(\boldsymbol{X}_2\ ) = \frac{2+2+0+1+1}{5} = 1.2$$

$$\overline{\mathrm{x}}_3 = \mathrm{E}(\boldsymbol{X}_3) = \frac{14+2+3+4+2}{5} = 5$$

Τα στοιχεία του πίνακα συνδιακύμανσης είναι:

$$s_{11} = \frac{1}{5-1}\sum_{i=1}^{5}(x_{i1} - \overline{\mathrm{x}}_1)^2$$
$$= \frac{(1-3.2)^2 + (8-3.2)^2 + (3-3.2)^2 + (4-3.2)^2 + (0-3.2)^2}{4} = \frac{38.8}{4} = 9.7$$

$$s_{22} = \frac{1}{5-1}\sum_{i=1}^{5}(x_{i2} - \overline{\mathrm{x}}_2)^2 = \frac{2.8}{4} = 0.7$$

$$s_{33} = \frac{1}{5-1}\sum_{i=1}^{5}(x_{i3} - \overline{\mathrm{x}}_3)^2 = \frac{104}{4} = 26$$

$$s_{12} = \frac{1}{5-1}\sum_{i=1}^{5}(x_{i1} - \overline{\mathrm{x}}_1)(x_{i2} - \overline{\mathrm{x}}_2) = \frac{2.8}{4} = 0.7$$

$$s_{13} = \frac{1}{5-1}\sum_{i=1}^{5}(x_{i1} - \overline{\mathrm{x}}_1)(x_{i3} - \overline{\mathrm{x}}_3) = \frac{-25}{4} = -6.25$$

$$s_{23} = \frac{1}{5-1}\sum_{i=1}^{5}(x_{i2} - \overline{\mathrm{x}}_2)(x_{i3} - \overline{\mathrm{x}}_3) = \frac{8}{4} = 2$$

Συνεπώς, ο πίνακας συνδιακύμανσης S_{LW} είναι:

$$S_{LW}=\begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix}=\begin{bmatrix} 9.7 & 0.7 & -6.25 \\ 0.7 & 0.7 & 2 \\ -6.25 & 2 & 26 \end{bmatrix}$$

ii. Σύμφωνα με το Πόρισμα 1.1 και τα στοιχεία του παραπάνω πίνακα συνδιακύμανσης S_{LW}, ο πίνακας συσχέτισης R_{LW} που προκύπτει είναι:

$$R_{LW}=(V^{1/2})^{-1}S_{LW}(V^{1/2})^{-1}$$

$$=\begin{bmatrix} \sqrt{9.7} & 0 & 0 \\ 0 & \sqrt{0.7} & 0 \\ 0 & 0 & \sqrt{26} \end{bmatrix}^{-1}\begin{bmatrix} 9.7 & 0.7 & -6.25 \\ 0.7 & 0.7 & 2 \\ -6.25 & 2 & 26 \end{bmatrix}\begin{bmatrix} \sqrt{9.7} & 0 & 0 \\ 0 & \sqrt{0.7} & 0 \\ 0 & 0 & \sqrt{26} \end{bmatrix}^{-1}$$

$$=\begin{bmatrix} 1 & 0.268 & -0.393 \\ 0.268 & 1 & 0.468 \\ -0.393 & 0.468 & 1 \end{bmatrix}$$

Προφανώς ο παραπάνω πίνακας R_{LW} είναι συμμετρικός και επαληθεύει τα αποτελέσματα της Πρότασης 1.9.

iii. Οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{LW} υπολογίζονται από την (1.3) από όπου έχουμε:

$$\det(S_{LW}-\lambda I)=0\Rightarrow -\lambda^3+36.4\lambda^2-233.637\lambda+80.156=0$$

Επομένως, οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{LW} είναι:

$$\lambda_1=0.363,\ \lambda_2=7.814,\ \lambda_3=28.222.$$

Από τις παραπάνω ιδιοτιμές παρατηρούμε ότι επαληθεύεται το (iii) στο Σχόλια 2.1 μια και οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{LW} είναι όλες θετικές $(\lambda_i \geq 0)$. ◊

Παράδειγμα 2.2

Έστω τρεις μεταβλητές $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$, οι οποίες έχουν τιμές όπως αυτές αναγράφονται στον Πίνακα 2.1.

Να βρεθούν:

i. ο πίνακας συνδιακύμανσης S_{LW},

ii. ο πίνακας συσχέτισης R_{LW}, και

iii. να υπολογιστούν οι ιδιοτιμές του S_{LW} .

$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$
1	_	_
3	1	2
_	_	1
8	2	_
0	_	3
5	7	8
2	_	1
3	_	3
4	1	4
0	_	2

Πίνακας 2.1: Σύνολο δεδομένων Παραδείγματος 2.2

Όπως και στο Παράδειγμα 2.1 εφαρμόζοντας τη μέθοδο παρατηρήσεων με ελλιπή δεδομένα στον παραπάνω πίνακα λαμβάνουμε υπόψη μόνο τις παρατηρήσεις που είναι πλήρεις. Έτσι, δημιουργείται το ακόλουθο σύνολο δεδομένων.

$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$
3	1	2
5	7	8
4	1	4

i. Ο πίνακας συνδιακύμανσης S_{LW} υπολογίζεται από την (1.18) του Ορισμού 1.18.

Οι μέσες τιμές κάθε μεταβλητής $\boldsymbol{X}_1, \boldsymbol{X}_2$ και $\boldsymbol{X}_3$ είναι:

$$\overline{\mathrm{x}}_1 = \mathrm{E}(\boldsymbol{X}_1\ ) = \frac{3+5+4}{3} = 4$$

$$\overline{\mathrm{x}}_2 = \mathrm{E}(\boldsymbol{X}_2\ ) = \frac{1+7+1}{3} = 3$$

$$\overline{\mathrm{x}}_3 = \mathrm{E}(\boldsymbol{X}_3) = \frac{2+8+4}{3} = 4.66666$$

Τα στοιχεία του πίνακα συνδιακύμανσης είναι:

$$s_{11} = \frac{1}{3-1}\sum_{i=1}^{3}(x_{i1} - \bar{x}_1)^2 = \frac{2}{2} = 1$$

$$s_{22} = \frac{1}{3-1}\sum_{i=1}^{5}(x_{i2} - \bar{x}_2)^2 = \frac{24}{2} = 12$$

$$s_{33} = \frac{1}{3-1}\sum_{i=1}^{3}(x_{i3} - \bar{x}_3)^2 = \frac{18.66666}{2} = 9.33333$$

$$s_{12} = \frac{1}{3-1}\sum_{i=1}^{3}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) = \frac{6}{2} = 3$$

$$s_{13} = \frac{1}{3-1}\sum_{i=1}^{3}(x_{i1} - \bar{x}_1)(x_{i3} - \bar{x}_3) = \frac{6}{2} = 3$$

$$s_{23} = \frac{1}{3-1}\sum_{i=1}^{3}(x_{i2} - \bar{x}_2)(x_{i3} - \bar{x}_3) = \frac{20}{2} = 10$$

Συνεπώς, ο πίνακας συνδιακύμανσης S_{LW} είναι:

$$S_{LW} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix} = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 12 & 10 \\ 3 & 10 & 9.33333 \end{bmatrix}$$

ii. Σύμφωνα με το Πόρισμα 1.1 και τα στοιχεία του παραπάνω πίνακα συνδιακύμανσης S_{LW}, ο πίνακας συσχέτισης R_{LW} που προκύπτει είναι:

$$\begin{aligned} R_{LW} &= (V^{1/2})^{-1} S_{LW} (V^{1/2})^{-1} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{12} & 0 \\ 0 & 0 & \sqrt{9.33333} \end{bmatrix}^{-1} \begin{bmatrix} 1 & 3 & 3 \\ 3 & 12 & 10 \\ 3 & 10 & 9.33333 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{12} & 0 \\ 0 & 0 & \sqrt{9.33333} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 0.866 & 0.982 \\ 0.866 & 1 & 0.9449 \\ 0.982 & 0.9449 & 1 \end{bmatrix} \end{aligned}$$

iii. Οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{LW} υπολογίζονται από την (1.3) και έχουμε:

$$\det(S_{LW} - \lambda I) = 0 \Rightarrow -\lambda^3 + 22.33333\lambda^2 - 15.33333\lambda = 0$$

Επομένως, οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{LW} είναι:

$$\lambda_1 = 0,\ \lambda_2 = 0.7091,\ \lambda_3 = 21.6242.$$

Σύμφωνα με την ιδιότητα (ii) της Πρότασης 1.3, ο πίνακας S_{LW} δεν αντιστρέφεται επειδή έχει το μηδέν ως ιδιοτιμή και είναι θετικά ημιορισμένος επαληθεύοντας την ιδιότητα (i) της Πρότασης 1.7.

◊

2.2. Μέθοδος EM (Expectation–maximization algorithm)

Ο αλγόριθμος EM είναι μια αποτελεσματική επαναληπτική διαδικασία για τον υπολογισμό εκτίμησης της μέγιστης πιθανοφάνειας (maximum likelihood), δηλαδή των πινάκων μέσης τιμής και συνδιακύμανσης [2]. Για την ακρίβεια, πρόκειται για μια μέθοδο βελτιστοποίησης της εκτίμησης κάποιων άγνωστων παραμέτρων ενός συνόλου δεδομένων [9], η οποία αποτελείται από δύο βήματα [15]:

a. το βήμα της πρόβλεψης (prediction step) και
b. το βήμα της εκτίμησης (estimation step).

Ο επαναληπτικός αλγόριθμος τερματίζει όταν ο πίνακας συνδιακύμανσης που προκύπτει κατά τη συγκεκριμένη επανάληψη ταυτίζεται με αυτόν της προηγούμενης.

Έστω ότι οι μεταβλητές $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ είναι ένα τυχαίο δείγμα από έναν κανονικό πληθυσμό, το οποίο έχει ελλιπή δεδομένα. ο αλγόριθμος, που κάνει πρόβλεψη και εκτίμηση, βασίζεται στην προσεγγιστική εκτίμηση των ελλιπών δεδομένων:

$$\boldsymbol{T}_1 = \sum_{i=1}^{n} \boldsymbol{X}_i = n\overline{\boldsymbol{x}} \tag{2.1}$$

και

$$T_2 = \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T = (n-1)S + n\overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^T \tag{2.2}$$

Ο αλγόριθμος εφαρμόζεται ως ακολούθως:

a. Κατά το στάδιο της πρόβλεψης: Αρχικά υποθέτουμε ότι η μέση τιμή $\tilde{\boldsymbol{\mu}}$ και ο πίνακας συνδιακύμανσης $\tilde{S}$ είναι άγνωστοι και πρέπει να υπολογιστούν. Στη συνέχεια, για κάθε παρατήρηση x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, p$, με ελλιπή δεδομένα, συμβολίζουμε με $\boldsymbol{x}_{ij}^{(1)}$ τα ελλιπή δεδομένα και με $\boldsymbol{x}_{ij}^{(2)}$ τα δεδομένα που είναι διαθέσιμα. Δηλαδή, $\boldsymbol{x}_{ij}^T = \left[\boldsymbol{x}_{ij}^{(1)} \quad \boldsymbol{x}_{ij}^{(2)}\right]$.

  Έχοντας ως δεδομένα τις εκτιμήσεις των μέσων τιμών $\tilde{\boldsymbol{\mu}}$ και του πίνακα συνδιακύμανσης $\tilde{S}$ που υπολογίστηκαν κατά τη διαδικασία της εκτίμησης, γίνεται χρήση της μέσης τιμής της υπό όρους κανονικής κατανομής του $\boldsymbol{x}_{ij}^{(1)}$, δεδομένου του $\boldsymbol{x}_{ij}^{(2)}$, προκειμένου να υπολογιστούν προσεγγιστικά τα ελλιπή δεδομένα. Για την ακρίβεια,

$$\tilde{\boldsymbol{x}}_{ij}^{(1)} = \tilde{\boldsymbol{\mu}}^{(1)} + \tilde{S}_{12}\tilde{S}_{12}^{-1}\left(\boldsymbol{x}_{ij}^{(2)} - \tilde{\boldsymbol{\mu}}^{(2)}\right) \tag{2.3}$$

εκτιμάει τη συμβολή του $\boldsymbol{x}_{ij}^{(1)}$ στο $\boldsymbol{T}_1$.

Η προβλεπόμενη συμβολή του $\boldsymbol{x}_{ij}^{(1)}$ στο T_2 είναι:

$$\widetilde{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(1)T}} = \tilde{S}_{11} + \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \tilde{\boldsymbol{x}}_{ij}^{(1)}\tilde{\boldsymbol{x}}_{ij}^{(1)T} \tag{2.4}$$

και

$$\widetilde{\boldsymbol{x}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(2)T}} = \tilde{\boldsymbol{x}}_{ij}^{(1)}\boldsymbol{x}_{ij}^{(2)T} \tag{2.5}$$

Οι συνεισφορές στις (2.3) και (2.4) αθροίζονται για όλα τα x_{ij} με ελλιπή δεδομένα. Τα αποτελέσματα συνδυάζονται με τα δεδομένα του δείγματος για να παραχθούν $\tilde{\boldsymbol{T}}_1$ και $\tilde{T}_2$.

b. Κατά το στάδιο της εκτίμησης υπολογίζεται:

$$\boldsymbol{\mu}_{EM} = \frac{\tilde{\boldsymbol{T}}_1}{n} \tag{2.6}$$

$$S_{EM} = \frac{1}{n}\tilde{T}_2 - \tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^T \tag{2.7}$$

Σχόλια 2.2

i. Τα κύρια πλεονεκτήματα του αλγορίθμου EM είναι:

   a. η απλότητά του, επειδή μπορεί να εκτελείται παράλληλα και οι απαιτήσεις του σε μνήμη τείνουν να είναι χαμηλές σε σύγκριση με άλλες μεθόδους,

   b. η ευκολία εφαρμογής του, γεγονός που συνεπάγεται ότι ο προγραμματισμός του αλγορίθμου είναι σε γενικές γραμμές μια εύκολη διαδικασία, η οποία τερματίζεται όταν τα στοιχεία των πινάκων $\boldsymbol{\mu}_{EM}$ και S_{EM} παραμείνουν ουσιαστικά αμετάβλητα.

   Το βασικό μειονέκτημα του είναι ότι σε κάποιες περιπτώσεις [5] η σύγκλισή του είναι αργή.

ii. Επειδή κατά την εφαρμογή της μεθόδου EM υπολογίστηκε ο πίνακας συνδιακύμανσης S_{EM}, σύμφωνα με το Πόρισμα 1.1 μπορεί να υπολογιστεί άμεσα και ο αντίστοιχος πίνακας συσχέτισης που θα συμβολίζεται R_{EM}.

iii. Σύμφωνα με τη θεωρία που αναπτύσσεται [15, chapter 4, 5], ο S_{EM} είναι **πάντα** ένας **θετικά ημιορισμένος** πίνακας.

Ο αντίστοιχος πίνακας συσχέτισης R_{EM} είναι **πάντα** ένας θετικά **ημιορισμένος πίνακας** (βλέπε Πρόταση 1.13). □

Χρησιμοποιώντας τον πίνακα με ελλιπή δεδομένα του Παραδείγματος 2.1 για τις τρεις μεταβλητές $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$, στο ακόλουθο παράδειγμα θα εφαρμόσουμε τη μέθοδο του αλγορίθμου ΕΜ προκειμένου να υπολογίσουμε τον πίνακα συνδιακύμανσης S_{EM} και εν συνεχεία τον πίνακα συσχέτισης R_{EM} .

Παράδειγμα 2.3

Χρησιμοποιώντας το σύνολο με τα ελλιπή δεδομένα του Παραδείγματος 2.1 και εφαρμόζοντας τον αλγόριθμο ΕΜ, να βρεθούν:

i. ο πίνακας συνδιακύμανσης S_{EM} ,
ii. ο πίνακας συσχέτισης R_{EM} , και
iii. να υπολογιστούν οι ιδιοτιμές του S_{EM} .

i. Ο πίνακας συνδιακύμανσης S_{EM} θα υπολογιστεί μετά από μια επανάληψη σύμφωνα με το (ib) από το Σχόλιο 2.2.

a. Αρχικά, κατά το στάδιο της πρόβλεψης υπολογίζουμε τις μέσες τιμές από τον τύπο (1.12) χρησιμοποιώντας τα διαθέσιμα στοιχεία ανά μεταβλητή:

$$\tilde{\mu}_1 = \frac{1+3+8+0+5+3+4+0}{8} = 3$$

$$\tilde{\mu}_2 = \frac{2+2+3+8+0+1+1}{7} = 2.4$$

$$\tilde{\mu}_3 = \frac{14+2+1+2+1+3+4+2}{8} = 3.6$$

Στη συνέχεια, αντικαθιστώντας την τιμή της μέσης τιμής σε κάθε ελλιπές δεδομένο, μπορούμε να υπολογίσουμε μια αρχική εκτίμηση του πίνακα συνδιακύμανσης με τη χρήση των (1.14) και (1.16) ως εξής:

$$\tilde{s}_{11} = \frac{1}{10}\sum_{i=1}^{10}(x_{i1} - \tilde{\mu}_1)^2$$
$$= \frac{(1-3)^2 + (3-3)^2 + (3-3)^2 + \ldots + (3-3)^2 + (4-3)^2 + (0-3)^2}{10} = 5.2$$

$$\tilde{s}_{22} = \frac{1}{10}\sum_{i=1}^{10}(x_{i2} - \tilde{\mu}_2)^2$$

$$= \frac{(2-2.4)^2 + (2.4-2.4)^2 + \cdots + (1-2.4)^2 + (1-2.4)^2}{10} = 4.2$$

$$\tilde{s}_{33} = \frac{1}{10}\sum_{i=1}^{10}(x_{i3} - \tilde{\mu}_3)^2$$

$$= \frac{(14-3.6)^2 + \cdots + (3.6-3.6)^2 + (1-3.6)^2 + \cdots + (2-3.6)^2}{10} = 12.98$$

$$\tilde{s}_{12} = \frac{1}{10}\sum_{i=1}^{10}(x_{i1} - \tilde{\mu}_1)(x_{i2} - \tilde{\mu}_2)$$

$$= \frac{(1-3)(2-2.4) + (3-3)(2.4-2.4) + \ \ldots \ + (0-3)(1-2.4)}{10} = -0.02$$

$$\tilde{s}_{13} = \frac{1}{10}\sum_{i=1}^{10}(x_{i1} - \tilde{\mu}_1)(x_{i3} - \tilde{\mu}_3)$$

$$= \frac{(1-3)(14-3.6) + \cdots + (0-3)(3.6-3.6) + \cdots + (0-3)(2-3.6)}{10} = -2.4$$

$$\tilde{s}_{23} = \frac{1}{10}\sum_{i=1}^{10}(x_{i2} - \tilde{\mu}_2)(x_{i3} - \tilde{\mu}_3)$$

$$= \frac{(2-2.4)(14-3.6) + \cdots + (5-2.4)(3.6-3.6) + \cdots + (0-2.4)(2-3.6)}{10} = -1.8$$

Συνεπώς, οι αρχικές εκτιμήσεις της μέσης τιμής και του πίνακα συνδιακύμανσης είναι:

$$\tilde{\boldsymbol{\mu}}_i = \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \\ \tilde{\mu}_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2.4 \\ 3.6 \end{bmatrix}$$

$$\tilde{S} = \begin{bmatrix} \tilde{s}_{11} & \tilde{s}_{12} & \tilde{s}_{13} \\ \tilde{s}_{21} & \tilde{s}_{22} & \tilde{s}_{23} \\ \tilde{s}_{31} & \tilde{s}_{32} & \tilde{s}_{33} \end{bmatrix} = \begin{bmatrix} 5.2 & -0.02 & -2.4 \\ -0.02 & 4.2 & -1.8 \\ -2.4 & -1.8 & 12.98 \end{bmatrix}$$

Το βήμα της πρόβλεψης συνίσταται από τη χρήση των αρχικών εκτιμήσεων $\tilde{\boldsymbol{\mu}}$ και $\tilde{S}$ προκειμένου να προβλεφθούν οι συνεισφορές των ελλιπών δεδομένων στα επαρκή στατιστικά στοιχεία $\boldsymbol{T}_1$ και T_2.

Παρατηρούμε ότι στο σύνολο δεδομένων που δίνεται στο παράδειγμα το x_{22} λείπει, έτσι διαμερίζουμε τους πίνακες $\tilde{\boldsymbol{\mu}}_i$ και $\tilde{S}$ ως εξής:

$$\tilde{\boldsymbol{\mu}}_i = \begin{bmatrix} \tilde{\mu}^{(1)} \\ \tilde{\mu}^{(2)} \end{bmatrix}, \ \tilde{S} = \begin{bmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ \tilde{S}_{21} & \tilde{S}_{22} \end{bmatrix}$$

όπου

$$\tilde{\boldsymbol{\mu}}^{(1)} = \left[\tilde{\mu}_1\right], \ \tilde{\boldsymbol{\mu}}^{(2)} = \begin{bmatrix} \tilde{\mu}_2 \\ \tilde{\mu}_3 \end{bmatrix}$$

$$\tilde{S}_{11} = \left[\tilde{s}_{11}\right], \ \tilde{S}_{12} = \left[\tilde{s}_{12} \quad \tilde{s}_{13}\right], \ \tilde{S}_{21} = \begin{bmatrix} \tilde{s}_{21} \\ \tilde{s}_{31} \end{bmatrix}, \ \tilde{S}_{22} = \begin{bmatrix} \tilde{s}_{22} & \tilde{s}_{23} \\ \tilde{s}_{32} & \tilde{s}_{33} \end{bmatrix}$$

και με τη χρήση των σχέσεων (2.3), (2.4) και (2.5) προβλέπουμε ότι:

$$\tilde{x}_{22} = \tilde{\mu}_2 + \tilde{S}_{12}\tilde{S}_{22}^{-1} \begin{bmatrix} x_{21} - \tilde{\mu}_1 \\ x_{23} - \tilde{\mu}_3 \end{bmatrix} = 2.4 + \left[-0.02 \quad -2.4\right] \begin{bmatrix} 4.2 & -1.8 \\ -1.8 & 12.98 \end{bmatrix}^{-1} \begin{bmatrix} 3-3 \\ 2-3.6 \end{bmatrix} = 2.7$$

$$\widetilde{x_{22}^2} = \tilde{s}_{22} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \tilde{x}_{22}^2 = 4.2 - \left[-0.02 \quad -2.4\right] \begin{bmatrix} 4.2 & -1.8 \\ -1.8 & 12.98 \end{bmatrix}^{-1} \begin{bmatrix} -0.02 \\ -2.4 \end{bmatrix} + 2.7^2 = 11$$

$$\widetilde{x_{22}\left[x_{21} \quad x_{23}\right]} = \tilde{x}_{22}\left[x_{21} \quad x_{23}\right] = 2.7\left[3 \quad 2\right] = \left[8.1 \quad 5.4\right]$$

Ομοίως, υπολογίζουμε τα x_{53} και x_{71}:

$$\tilde{x}_{53} = \tilde{\mu}_3 + \tilde{S}_{12}\tilde{S}_{22}^{-1} \begin{bmatrix} x_{51} - \tilde{\mu}_1 \\ x_{52} - \tilde{\mu}_2 \end{bmatrix} = 3.6 + \left[-0.02 \quad -2.4\right] \begin{bmatrix} 4.2 & -1.8 \\ -1.8 & 12.98 \end{bmatrix}^{-1} \begin{bmatrix} 0-3 \\ 3-2.4 \end{bmatrix} = 3.75$$

$$\widetilde{x_{53}^2} = \tilde{s}_{33} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \tilde{x}_{22}^2 = 12.98 - \left[-0.02 \quad -2.4\right] \begin{bmatrix} 4.2 & -1.8 \\ -1.8 & 12.98 \end{bmatrix}^{-1} \begin{bmatrix} -0.02 \\ -2.4 \end{bmatrix} + 3.75^2 = 26.5$$

$$\widetilde{x_{53}\left[x_{51} \quad x_{52}\right]} = \tilde{x}_{53}\left[x_{51} \quad x_{52}\right] = 3.75\left[0 \quad 3\right] = \left[0 \quad 11.25\right]$$

$$\tilde{x}_{71} = \tilde{\mu}_1 + \tilde{S}_{12}\tilde{S}_{22}^{-1} \begin{bmatrix} x_{72} - \tilde{\mu}_2 \\ x_{73} - \tilde{\mu}_3 \end{bmatrix} = 3 + \left[-0.02 \quad -2.4\right] \begin{bmatrix} 4.2 & -1.8 \\ -1.8 & 12.98 \end{bmatrix}^{-1} \begin{bmatrix} 8-2.4 \\ 1-3.6 \end{bmatrix} = 3$$

$$\widetilde{x_{71}^2} = \tilde{s}_{11} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \tilde{x}_{71}^2 = 5.2 - \left[-0.02 \quad -2.4\right] \begin{bmatrix} 4.2 & -1.8 \\ -1.8 & 12.98 \end{bmatrix}^{-1} \begin{bmatrix} -0.02 \\ -2.4 \end{bmatrix} + 3^2 = 13.7$$

$$\widetilde{x_{71}\left[x_{72} \quad x_{73}\right]} = \tilde{x}_{71}\left[x_{72} \quad x_{73}\right] = 3\left[8 \quad 1\right] = \left[24 \quad 3\right]$$

Στην περίπτωση που λείπουν δυο δεδομένα σε μια παρατήρηση, όπως τα x_{31} και x_{32}, οι διαμερίσεις των πινάκων $\tilde{\boldsymbol{\mu}}_i$ και $\tilde{S}$ είναι οι ακόλουθες

$$\tilde{\boldsymbol{\mu}}_i = \begin{bmatrix} \tilde{\boldsymbol{\mu}}^{(1)} \\ \tilde{\boldsymbol{\mu}}^{(2)} \end{bmatrix}, \qquad \tilde{S} = \begin{bmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ \tilde{S}_{21} & \tilde{S}_{22} \end{bmatrix}$$

όπου

$$\tilde{\boldsymbol{\mu}}^{(1)} = \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{bmatrix}, \qquad \tilde{\boldsymbol{\mu}}^{(2)} = \begin{bmatrix} \tilde{\mu}_3 \end{bmatrix}$$

$$\tilde{S}_{11} = \begin{bmatrix} \tilde{s}_{11} & \tilde{s}_{12} \\ \tilde{s}_{21} & \tilde{s}_{22} \end{bmatrix}, \quad \tilde{S}_{12} = \begin{bmatrix} \tilde{s}_{13} & \tilde{s}_{23} \end{bmatrix}, \quad \tilde{S}_{21} = \begin{bmatrix} \tilde{s}_{31} \\ \tilde{s}_{32} \end{bmatrix}, \quad \tilde{S}_{22} = \begin{bmatrix} \tilde{s}_{33} \end{bmatrix}$$

και με τη χρήση της σχέσης (2.3) προβλέπουμε ότι:

$$\widetilde{\begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix}} = \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{bmatrix} + \tilde{S}_{12}\tilde{S}_{22}^{-1}\left[x_{33} - \tilde{\mu}_3\right] = \begin{bmatrix} 3 \\ 2.4 \end{bmatrix} + \begin{bmatrix} -2.4 \\ -1.8 \end{bmatrix} [12.98]^{-1} [1 - 3.6] = \begin{bmatrix} 3.5 \\ 2.8 \end{bmatrix}$$

$$\widetilde{\begin{bmatrix} x_{62} \\ x_{63} \end{bmatrix}} = \begin{bmatrix} \tilde{\mu}_2 \\ \tilde{\mu}_3 \end{bmatrix} + \tilde{S}_{12}\tilde{S}_{22}^{-1}\left[x_{11} - \tilde{\mu}_1\right] = \begin{bmatrix} 2.4 \\ 3.6 \end{bmatrix} + \begin{bmatrix} -2.4 \\ -1.8 \end{bmatrix} [12.98]^{-1} [5 - 3] = \begin{bmatrix} 2 \\ 3.3 \end{bmatrix}$$

για τη συμβολή του $\boldsymbol{T}_1$. Επίσης, από τον τύπο (2.4) υπολογίζουμε τη συμβολή του T_2:

$$\begin{bmatrix} \widetilde{x_{31}^2} & \widetilde{x_{31}x_{32}} \\ \widetilde{x_{31}x_{32}} & \widetilde{x_{32}^2} \end{bmatrix} = \tilde{S}_{11} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix} \begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix}^T$$

$$= \begin{bmatrix} 5.2 & -0.02 \\ -0.02 & 4.2 \end{bmatrix} - \begin{bmatrix} -2.4 \\ -1.8 \end{bmatrix} [12.98]^{-1} \begin{bmatrix} -2.4 & -1.8 \end{bmatrix} + \begin{bmatrix} 3.5 \\ 2.8 \end{bmatrix} \begin{bmatrix} 3.5 \\ 2.8 \end{bmatrix}^T$$

$$= \begin{bmatrix} 17 & 9.4 \\ 9.4 & 11.8 \end{bmatrix}$$

$$\widetilde{\begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix} (x_{33})} = \widetilde{\begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix}} (x_{33}) = \begin{bmatrix} 3.5 \\ 2.8 \end{bmatrix}$$

$$\begin{bmatrix} \widetilde{x_{62}^2} & \widetilde{x_{62}x_{63}} \\ \widetilde{x_{62}x_{63}} & \widetilde{x_{63}^2} \end{bmatrix} = \tilde{S}_{11} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \begin{bmatrix} x_{62} \\ x_{63} \end{bmatrix} \begin{bmatrix} x_{62} \\ x_{63} \end{bmatrix}^T$$

$$= \begin{bmatrix} 5.2 & -0.02 \\ -0.02 & 4.2 \end{bmatrix} - \begin{bmatrix} -2.4 \\ -1.8 \end{bmatrix} [12.98]^{-1} \begin{bmatrix} -2.4 & -1.8 \end{bmatrix} + \begin{bmatrix} 2 \\ 3.3 \end{bmatrix} \begin{bmatrix} 2 \\ 3.3 \end{bmatrix}^T$$

$$= \begin{bmatrix} 8.8 & 6.2 \\ 6.2 & 14.8 \end{bmatrix}$$

$$\widetilde{\begin{bmatrix} x_{62} \\ x_{63} \end{bmatrix}(x_{61})} = \widetilde{\begin{bmatrix} x_{62} \\ x_{63} \end{bmatrix}}(x_{61}) = \begin{bmatrix} 2 \\ 3.3 \end{bmatrix}(5) = \begin{bmatrix} 10 \\ 16.5 \end{bmatrix}$$

Άρα, σύμφωνα με τη (2.1)

$$\tilde{\boldsymbol{T}}_1 = \begin{bmatrix} x_{11} + x_{21} + \tilde{x}_{31} + x_{41} + x_{51} + x_{61} + \tilde{x}_{71} + x_{81} + x_{91} + x_{101} \\ x_{12} + \tilde{x}_{22} + \tilde{x}_{32} + x_{42} + x_{52} + \tilde{x}_{62} + x_{72} + x_{82} + x_{92} + x_{102} \\ x_{13} + x_{23} + x_{33} + x_{43} + \tilde{x}_{53} + \tilde{x}_{63} + x_{73} + x_{83} + x_{93} + x_{103} \end{bmatrix}$$

$$= \begin{bmatrix} 1+3+3.5+8+0+5+3+3+4+0 \\ 2+2.7+2.8+2+3+2+8+0+1+1 \\ 14+2+1+2+3.75+3.3+1+3+4+2 \end{bmatrix} = \begin{bmatrix} 30.5 \\ 24.5 \\ 36 \end{bmatrix}$$

$$\tilde{T}_2 = \begin{bmatrix} x_{11}^2 + x_{21}^2 + \widetilde{x_{31}^2} + \ldots + \widetilde{x_{71}^2} + x_{81}^2 + x_{91}^2 + x_{101}^2 & & \\ x_{11}x_{12} + \widetilde{x_{21}x_{22}} + \widetilde{x_{31}x_{32}} + \ldots + x_{101}x_{102} & x_{12}^2 + \widetilde{x_{22}^2} + \widetilde{x_{32}^2} + \ldots + x_{102}^2 & \\ x_{11}x_{13} + x_{21}x_{23} + \widetilde{x_{31}x_{33}} + \ldots + x_{101}x_{103} & x_{12}x_{13} + x_{22}x_{23} + \ldots + x_{102}x_{103} & x_{13}^2 + \ldots + \widetilde{x_{53}^2} + \ldots + x_{103}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 154.7 & 73.5 & 84 \\ 73.5 & 114.6 & 71.65 \\ 84 & 71.65 & 276.3 \end{bmatrix}$$

Σε αυτό το σημείο ολοκληρώνεται η διαδικασία πρόβλεψης για την 1^{η} επανάληψη.

b. Η διαδικασία της εκτίμησης με τη χρήση των σχέσεων (2.6) και (2.7) παρέχει τις εκτιμήσεις:

$$\boldsymbol{\mu}_{EM} = \frac{\tilde{\boldsymbol{T}}_1}{n} = \frac{1}{10}\begin{bmatrix} 30.5 \\ 24.5 \\ 36 \end{bmatrix} = \begin{bmatrix} 3.05 \\ 2.45 \\ 3.6 \end{bmatrix}$$

$$S_{EM} = \frac{1}{n}\tilde{T}_2 - \tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^{\mathrm{T}} = \frac{1}{10}\begin{bmatrix} 154.7 & 73.5 & 84 \\ 73.5 & 114.6 & 71.65 \\ 84 & 71.65 & 276.3 \end{bmatrix} - \begin{bmatrix} 3.05 \\ 2.45 \\ 3.6 \end{bmatrix}\begin{bmatrix} 3.05 \\ 2.45 \\ 3.6 \end{bmatrix}^T$$

$$= \begin{bmatrix} 6.16 & -0.13 & -2.58 \\ -0.13 & 5.45 & -1.65 \\ -2.58 & -1.65 & 14.67 \end{bmatrix}$$

Παρατηρώ ότι τα στοιχεία της διαγωνίου του παραπάνω πίνακα S_{EM} είναι μεγαλύτερα συγκριτικά με τα στοιχεία του πίνακα συνδιακύμανσης $\tilde{S}$ που υπολογίστηκαν κατά την αρχική εκτίμηση.

ii. Σύμφωνα με το Πόρισμα 1.1 και τα στοιχεία του παραπάνω πίνακα συνδιακύμανσης S_{EM} από το (ib), ο πίνακας συσχέτισης R_{EM} που προκύπτει είναι:

$$\begin{aligned} R_{EM} &= (V^{1/2})^{-1} S_{EM} (V^{1/2})^{-1} \\ &= \begin{bmatrix} 6.16 & 0 & 0 \\ 0 & 5.45 & 0 \\ 0 & 0 & 14.67 \end{bmatrix}^{-1} \begin{bmatrix} 6.16 & -0.13 & -2.58 \\ -0.13 & 5.45 & -1.65 \\ -2.58 & -1.65 & 14.67 \end{bmatrix} \begin{bmatrix} 6.16 & 0 & 0 \\ 0 & 5.45 & 0 \\ 0 & 0 & 14.67 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & -0.02 & -0.27 \\ -0.02 & 1 & -0.18 \\ -0.27 & -0.18 & 1 \end{bmatrix} \end{aligned}$$

iii. Οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{EM} υπολογίζονται από την (1.3) από όπου έχουμε:

$$\det(S_{EM} - \lambda I) = 0 \Rightarrow -\lambda^3 + 26.28\lambda^2 - 194.495\lambda + 438.098 = 0$$

Επομένως, οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{EM} είναι:

$$\lambda_1 = 4.7515,\ \lambda_2 = 5.899,\ \lambda_3 = 15.629.$$

Από τα παραπάνω είναι φανερό ότι ο πίνακας S_{EM} είναι θετικά ημιορισμένος επαληθεύοντας το (iii) στο Σχόλιο 2.2.

◊

Παράδειγμα 2.4

Χρησιμοποιώντας το σύνολο με τα ελλιπή δεδομένα του Παραδείγματος 2.2 και εφαρμόζοντας τον αλγόριθμο ΕΜ, να βρεθούν:

i. ο πίνακας συνδιακύμανσης S_{EM} ,

ii. ο πίνακας συσχέτισης R_{EM} , και

iii. να υπολογιστούν οι ιδιοτιμές του S_{EM} .

ii. Ο πίνακας συνδιακύμανσης S_{EM} θα υπολογιστεί μετά από μια επανάληψη σύμφωνα με το (ib) από το Σχόλιο 2.2.

a. Αρχικά, κατά το στάδιο της πρόβλεψης υπολογίζουμε τις μέσες τιμές από τον τύπο (1.12) χρησιμοποιώντας τα διαθέσιμα στοιχεία ανά μεταβλητή:

$$\tilde{\mu}_1 = \frac{1+3+8+0+5+2+3+4+0}{9} = \frac{26}{9} = 2.88$$

$$\tilde{\mu}_2 = \frac{1+2+7+1}{4} = \frac{11}{4} = 2.75$$

$$\tilde{\mu}_3 = \frac{2+1+3+8+1+3+4+2}{8} = \frac{24}{8} = 3$$

Στη συνέχεια, αντικαθιστώντας την τιμή της μέσης τιμής σε κάθε ελλιπές δεδομένο, μπορούμε να υπολογίσουμε μια αρχική εκτίμηση του πίνακα συνδιακύμανσης με τη χρήση των (1.14) και (1.16) ως εξής:

$$\tilde{s}_{11} = \frac{1}{10}\sum_{i=1}^{10}(x_{i1} - \tilde{\mu}_1)^2 = \frac{52.8}{10} = 5.28$$

$$\tilde{s}_{22} = \frac{1}{10}\sum_{i=1}^{10}(x_{i2} - \tilde{\mu}_2)^2 = \frac{24.7}{10} = 2.47$$

$$\tilde{s}_{33} = \frac{1}{10}\sum_{i=1}^{10}(x_{i3} - \tilde{\mu}_3)^2 = \frac{36}{10} = 3.6$$

$$\tilde{s}_{12} = \frac{1}{10}\sum_{i=1}^{10}(x_{i1} - \tilde{\mu}_1)(x_{i2} - \tilde{\mu}_2) = \frac{3}{10} = 0.3$$

$$\tilde{s}_{13} = \frac{1}{10}\sum_{i=1}^{10}(x_{i1} - \tilde{\mu}_1)(x_{i3} - \tilde{\mu}_3) = \frac{16}{10} = 1.6$$

$$\tilde{s}_{23} = \frac{1}{10}\sum_{i=1}^{10}(x_{i2} - \tilde{\mu}_2)(x_{i3} - \tilde{\mu}_3) = \frac{21}{10} = 2.1$$

Συνεπώς, οι αρχικές εκτιμήσεις της μέσης τιμής και του πίνακα συνδιακύμανσης είναι:

$$\tilde{\boldsymbol{\mu}}_i = \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \\ \tilde{\mu}_3 \end{bmatrix} = \begin{bmatrix} 2.88 \\ 2.75 \\ 3 \end{bmatrix}$$

$$\tilde{S} = \begin{bmatrix} \tilde{s}_{11} & \tilde{s}_{12} & \tilde{s}_{13} \\ \tilde{s}_{21} & \tilde{s}_{22} & \tilde{s}_{23} \\ \tilde{s}_{31} & \tilde{s}_{32} & \tilde{s}_{33} \end{bmatrix} = \begin{bmatrix} 5.28 & 0.3 & 1.6 \\ 0.3 & 2.47 & 2.1 \\ 1.6 & 2.1 & 3.6 \end{bmatrix}$$

Το βήμα της πρόβλεψης συνίσταται από τη χρήση των αρχικών εκτιμήσεων $\tilde{\boldsymbol{\mu}}$ και $\tilde{S}$ προκειμένου να προβλεφθούν οι συνεισφορές των ελλιπών δεδομένων στα επαρκή στατιστικά στοιχεία $\boldsymbol{T}_1$ και T_2.

$$\tilde{x}_{43} = \tilde{\mu}_3 + \tilde{S}_{12}\tilde{S}_{22}^{-1} \begin{bmatrix} x_{41} - \tilde{\mu}_1 \\ x_{42} - \tilde{\mu}_2 \end{bmatrix} = 3 + \begin{bmatrix} 0.3 & 1.6 \end{bmatrix} \begin{bmatrix} 2.47 & 2.1 \\ 2.1 & 3.6 \end{bmatrix}^{-1} \begin{bmatrix} 8-2.88 \\ 2-2.75 \end{bmatrix} = -0.16$$

$$\widetilde{x_{43}^{2}} = \tilde{s}_{22} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \tilde{x}_{22}^{2} = 3.6 - \begin{bmatrix} 0.3 & 1.6 \end{bmatrix} \begin{bmatrix} 2.47 & 2.1 \\ 2.1 & 3.6 \end{bmatrix}^{-1} \begin{bmatrix} 0.3 \\ 1.6 \end{bmatrix} + (-0.16)^2 = 2.59$$

$$\widetilde{x_{43}\begin{bmatrix} x_{41} & x_{42} \end{bmatrix}} = \tilde{x}_{43}\begin{bmatrix} x_{41} & x_{42} \end{bmatrix} = -0.16\begin{bmatrix} 8 & 2 \end{bmatrix} = \begin{bmatrix} -1.28 & -0.32 \end{bmatrix}$$

$$\tilde{x}_{52} = \tilde{\mu}_2 + \tilde{S}_{12}\tilde{S}_{22}^{-1} \begin{bmatrix} x_{51} - \tilde{\mu}_1 \\ x_{53} - \tilde{\mu}_3 \end{bmatrix} = 2.75 + \begin{bmatrix} 0.3 & 1.6 \end{bmatrix} \begin{bmatrix} 2.47 & 2.1 \\ 2.1 & 3.6 \end{bmatrix}^{-1} \begin{bmatrix} 0-2.88 \\ 3-3 \end{bmatrix} = 4.21$$

$$\widetilde{x_{52}^{2}} = \tilde{s}_{22} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \tilde{x}_{52}^{2} = 2.47 - \begin{bmatrix} 0.3 & 1.6 \end{bmatrix} \begin{bmatrix} 2.47 & 2.1 \\ 2.1 & 3.6 \end{bmatrix}^{-1} \begin{bmatrix} 0.3 \\ 1.6 \end{bmatrix} + 4.21^2 = 19.16$$

$$\widetilde{x_{52}\begin{bmatrix} x_{51} & x_{53} \end{bmatrix}} = \tilde{x}_{52}\begin{bmatrix} x_{51} & x_{53} \end{bmatrix} = 4.21\begin{bmatrix} 0 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 12.63 \end{bmatrix}$$

$$\tilde{x}_{72} = \tilde{\mu}_2 + \tilde{S}_{12}\tilde{S}_{22}^{-1} \begin{bmatrix} x_{71} - \tilde{\mu}_1 \\ x_{73} - \tilde{\mu}_3 \end{bmatrix} = 2.75 + \begin{bmatrix} 0.3 & 1.6 \end{bmatrix} \begin{bmatrix} 2.47 & 2.1 \\ 2.1 & 3.6 \end{bmatrix}^{-1} \begin{bmatrix} 2-2.88 \\ 1-3 \end{bmatrix} = 1.7$$

$$\widetilde{x_{72}^{2}} = \tilde{s}_{22} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \tilde{x}_{72}^{2} = 2.47 - \begin{bmatrix} 0.3 & 1.6 \end{bmatrix} \begin{bmatrix} 2.47 & 2.1 \\ 2.1 & 3.6 \end{bmatrix}^{-1} \begin{bmatrix} 0.3 \\ 1.6 \end{bmatrix} + 1.7^2 = 4.32$$

$$\widetilde{x_{72}\begin{bmatrix} x_{71} & x_{73} \end{bmatrix}} = \tilde{x}_{72}\begin{bmatrix} x_{71} & x_{73} \end{bmatrix} = 1.7\begin{bmatrix} 2 & 1 \end{bmatrix} = \begin{bmatrix} 3.4 & 1.7 \end{bmatrix}$$

$$\tilde{x}_{82} = \tilde{\mu}_2 + \tilde{S}_{12}\tilde{S}_{22}^{-1} \begin{bmatrix} x_{81} - \tilde{\mu}_1 \\ x_{83} - \tilde{\mu}_3 \end{bmatrix} = 2.75 + \begin{bmatrix} 0.3 & 1.6 \end{bmatrix} \begin{bmatrix} 2.47 & 2.1 \\ 2.1 & 3.6 \end{bmatrix}^{-1} \begin{bmatrix} 3-2.88 \\ 3-3 \end{bmatrix} = 2.69$$

$$\widetilde{x_{82}^{2}} = \tilde{s}_{22} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \tilde{x}_{82}^{2} = 2.47 - \begin{bmatrix} 0.3 & 1.6 \end{bmatrix} \begin{bmatrix} 2.47 & 2.1 \\ 2.1 & 3.6 \end{bmatrix}^{-1} \begin{bmatrix} 0.3 \\ 1.6 \end{bmatrix} + 2.69^2 = 8.67$$

$$\widetilde{x_{82}\begin{bmatrix} x_{81} & x_{83} \end{bmatrix}} = \tilde{x}_{82}\begin{bmatrix} x_{81} & x_{83} \end{bmatrix} = 2.69\begin{bmatrix} 3 & 3 \end{bmatrix} = \begin{bmatrix} 8.07 & 8.07 \end{bmatrix}$$

$$\tilde{x}_{102} = \tilde{\mu}_2 + \tilde{S}_{12}\tilde{S}_{22}^{-1}\begin{bmatrix} x_{101}-\tilde{\mu}_1 \\ x_{103}-\tilde{\mu}_3 \end{bmatrix} = 2.75 + \begin{bmatrix} 0.3 & 1.6 \end{bmatrix}\begin{bmatrix} 2.47 & 2.1 \\ 2.1 & 3.6 \end{bmatrix}^{-1}\begin{bmatrix} 0-2.88 \\ 2-3 \end{bmatrix} = 3.47$$

$$\widetilde{x_{102}^2} = \tilde{s}_{22} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \tilde{x}_{102}^2 = 2.47 - \begin{bmatrix} 0.3 & 1.6 \end{bmatrix}\begin{bmatrix} 2.47 & 2.1 \\ 2.1 & 3.6 \end{bmatrix}^{-1}\begin{bmatrix} 0.3 \\ 1.6 \end{bmatrix} + 3.47^2 = 13.48$$

$$\widetilde{x_{102}\begin{bmatrix} x_{101} & x_{103} \end{bmatrix}} = \tilde{x}_{102}\begin{bmatrix} x_{101} & x_{7103} \end{bmatrix} = 3.47\begin{bmatrix} 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 6.94 \end{bmatrix}$$

$$\widetilde{\begin{bmatrix} x_{12} \\ x_{13} \end{bmatrix}} = \begin{bmatrix} \tilde{\mu}_2 \\ \tilde{\mu}_3 \end{bmatrix} + \tilde{S}_{12}\tilde{S}_{22}^{-1}\left[x_{11}-\tilde{\mu}_1\right] = \begin{bmatrix} 2.75 \\ 3 \end{bmatrix} + \begin{bmatrix} 1.6 \\ 2.1 \end{bmatrix}\begin{bmatrix} 3.6 \end{bmatrix}^{-1}\begin{bmatrix} 1-2.88 \end{bmatrix} = \begin{bmatrix} 1.91 \\ 1.9 \end{bmatrix}$$

$$\begin{bmatrix} \widetilde{x_{12}^2} & \widetilde{x_{12}x_{13}} \\ \widetilde{x_{12}x_{13}} & \widetilde{x_{13}^2} \end{bmatrix} = \tilde{S}_{11} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \begin{bmatrix} x_{12} \\ x_{13} \end{bmatrix}\begin{bmatrix} x_{12} \\ x_{13} \end{bmatrix}^T$$

$$= \begin{bmatrix} 5.28 & 0.3 \\ 0.3 & 2.47 \end{bmatrix} - \begin{bmatrix} 1.6 \\ 2.1 \end{bmatrix}\begin{bmatrix} 3.6 \end{bmatrix}^{-1}\begin{bmatrix} 1.6 & 2.1 \end{bmatrix} + \begin{bmatrix} 1.91 \\ 1.9 \end{bmatrix}\begin{bmatrix} 1.91 \\ 1.9 \end{bmatrix}^T$$

$$= \begin{bmatrix} 8.2 & 2.99 \\ 2.99 & 4.85 \end{bmatrix}$$

$$\widetilde{\begin{bmatrix} x_{12} \\ x_{13} \end{bmatrix}(x_{11})} = \widetilde{\begin{bmatrix} x_{12} \\ x_{13} \end{bmatrix}}(x_{11}) = \begin{bmatrix} 1.91 \\ 1.9 \end{bmatrix}$$

$$\widetilde{\begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix}} = \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{bmatrix} + \tilde{S}_{12}\tilde{S}_{22}^{-1}\left[x_{33}-\tilde{\mu}_3\right] = \begin{bmatrix} 2.88 \\ 2.75 \end{bmatrix} + \begin{bmatrix} 1.6 \\ 2.1 \end{bmatrix}\begin{bmatrix} 3.6 \end{bmatrix}^{-1}\begin{bmatrix} 1-3 \end{bmatrix} = \begin{bmatrix} 1.99 \\ 1.58 \end{bmatrix}$$

$$\begin{bmatrix} \widetilde{x_{31}^2} & \widetilde{x_{31}x_{32}} \\ \widetilde{x_{31}x_{32}} & \widetilde{x_{32}^2} \end{bmatrix} = \tilde{S}_{11} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} + \begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix}\begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix}^T$$

$$= \begin{bmatrix} 5.28 & 0.3 \\ 0.3 & 2.47 \end{bmatrix} - \begin{bmatrix} 1.6 \\ 2.1 \end{bmatrix}\begin{bmatrix} 3.6 \end{bmatrix}^{-1}\begin{bmatrix} 1.6 & 2.1 \end{bmatrix} + \begin{bmatrix} 1.99 \\ 1.58 \end{bmatrix}\begin{bmatrix} 1.99 \\ 1.58 \end{bmatrix}^T$$

$$= \begin{bmatrix} 8.53 & 2.5 \\ 2.5 & 3.74 \end{bmatrix}$$

$$\widetilde{\begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix}(x_{33})} = \widetilde{\begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix}}(x_{33}) = \begin{bmatrix} 1.99 \\ 1.58 \end{bmatrix}$$

Άρα , σύμφωνα με τη (2.1)

$$\tilde{\boldsymbol{T}}_1 = \begin{bmatrix} x_{11} + x_{21} + \tilde{x}_{31} + x_{41} + x_{51} + x_{61} + x_{71} + x_{81} + x_{91} + x_{101} \\ \tilde{x}_{12} + x_{22} + \tilde{x}_{32} + x_{42} + \tilde{x}_{52} + x_{62} + \tilde{x}_{72} + \tilde{x}_{82} + x_{92} + \tilde{x}_{102} \\ \tilde{x}_{13} + x_{23} + x_{33} + \tilde{x}_{43} + x_{53} + x_{63} + x_{73} + x_{83} + x_{93} + x_{103} \end{bmatrix}$$

$$= \begin{bmatrix} 1+3+1.99+8+0+5+2+3+4+0 \\ 1.91+1+1.58+2+4.21+7+1.7+2.69+1+3.47 \\ 1.99+2+1-0.16+3+8+1+3+4+2 \end{bmatrix} = \begin{bmatrix} 27.99 \\ 26.56 \\ 25.8 \end{bmatrix}$$

$$\tilde{T}_2 = \begin{bmatrix} x_{11}^2 + x_{21}^2 + \widetilde{x_{31}^2} + \ldots + x_{71}^2 + x_{81}^2 + x_{91}^2 + x_{101}^2 & & \\ \widetilde{x_{11}x_{12}} + x_{21}x_{22} + \widetilde{x_{31}x_{32}} + \ldots + \widetilde{x_{101}x_{102}} & \widetilde{x_{12}^2} + x_{22}^2 + \widetilde{x_{32}^2} + \ldots + \widetilde{x_{102}^2} & \\ \widetilde{x_{11}x_{13}} + x_{21}x_{23} + \widetilde{x_{31}x_{33}} + \ldots + x_{101}x_{103} & \widetilde{x_{12}x_{13}} + x_{22}x_{23} + \ldots + \widetilde{x_{102}x_{103}} & \widetilde{x_{13}^2} + \ldots + x_{53}^2 + \ldots + x_{103}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 136.5 & 73.8 & 75.6 \\ 73.8 & 112.6 & 95.6 \\ 75.6 & 95.6 & 115.4 \end{bmatrix}$$

Σε αυτό το σημείο ολοκληρώνεται η διαδικασία πρόβλεψης για την 1^{η} επανάληψη.

b. Η διαδικασία της εκτίμησης με τη χρήση των σχέσεων (2.6) και (2.7) παρέχει τις εκτιμήσεις:

$$\boldsymbol{\mu}_{EM} = \frac{\tilde{\boldsymbol{T}}_1}{n} = \frac{1}{10} \begin{bmatrix} 27.99 \\ 26.56 \\ 25.8 \end{bmatrix} = \begin{bmatrix} 2.799 \\ 2.656 \\ 2.58 \end{bmatrix}$$

$$S_{EM} = \frac{1}{n}\tilde{T}_2 - \tilde{\mu}\tilde{\mu}^{\mathrm{T}} = \frac{1}{10} \begin{bmatrix} 136.5 & 73.8 & 75.6 \\ 73.8 & 112.6 & 95.6 \\ 75.6 & 95.6 & 115.4 \end{bmatrix} - \begin{bmatrix} 2.799 \\ 2.656 \\ 2.58 \end{bmatrix} \begin{bmatrix} 2.799 \\ 2.656 \\ 2.58 \end{bmatrix}^T$$

$$= \begin{bmatrix} 5.8156 & -0.0541 & 0.3386 \\ -0.0541 & 4.2057 & 2.7075 \\ 0.3386 & 2.7075 & 4.8836 \end{bmatrix}$$

ii. Σύμφωνα με το Πόρισμα 1.1 και τα στοιχεία του παραπάνω πίνακα συνδιακύμανσης S_{EM} από το (ib), ο πίνακας συσχέτισης R_{EM} που προκύπτει είναι:

$$R_{EM} = (V^{1/2})^{-1} S_{EM} (V^{1/2})^{-1}$$

$$= \begin{bmatrix} \sqrt{5.8156} & 0 & 0 \\ 0 & \sqrt{4.2057} & 0 \\ 0 & 0 & \sqrt{4.8836} \end{bmatrix}^{-1} \begin{bmatrix} 5.8156 & -0.0541 & 0.3386 \\ -0.0541 & 4.2057 & 2.7075 \\ 0.3386 & 2.7075 & 4.8836 \end{bmatrix} (V^{1/2})^{-1}$$

$$= \begin{bmatrix} 1 & -0.0109 & 0.0635 \\ -0.0109 & 1 & 0.5974 \\ 0.0635 & 0.5974 & 1 \end{bmatrix}$$

iii. Οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{EM} υπολογίζονται από την (1.3) από όπου έχουμε:

$$\det(S_{EM} - \lambda I) = 0 \Rightarrow -\lambda^3 + 14.9049\lambda^2 - 65.9501\lambda + 76.2174 = 0$$

Επομένως, οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{EM} είναι:

$$\lambda_1 = 1.7985,\ \lambda_2 = 5.8009,\ \lambda_3 = 7.3055.$$

Από τα παραπάνω είναι φανερό ότι ο πίνακας S_{EM} είναι θετικά ημιορισμένος επαληθεύοντας το (iii) στο Σχόλιο 2.2. ◊

2.3. Μέθοδος διαγραφής ανά ζεύγη

Μια άλλη μέθοδος υπολογισμού του πίνακα συνδιακύμανσης, ο οποίος προέρχεται από ένα σύνολο με κάποια ελλιπή δεδομένα, είναι η μέθοδος διαγραφής ανά ζεύγη. Αυτή η μέθοδος αξιοποιεί όλα τα διαθέσιμα δεδομένα του συνόλου [3, 26], δηλαδή χρησιμοποιεί ακόμα και τις παρατηρήσεις που περιέχουν κάποιες ελλιπείς τιμές, αρκεί να μπορεί να γίνει η ανάλυση κατά ζεύγη (δυο) μεταβλητών όπως περιγράφεται ακολούθως. Έτσι, το βασικό χαρακτηριστικό της είναι ότι κατά την ανάλυση δυο μεταβλητών μια παρατήρηση, που δεν έχει κάποια(ες) διαθέσιμη(ες) τιμή(ες), μπορεί να εξαιρεθεί και η ίδια παρατήρηση μπορεί να συμμετέχει κατά την ανάλυση δυο διαφορετικών μεταβλητών. Εν κατακλείδι η μέθοδος διαγραφής ανά ζεύγη υπολογίζει πίνακα συνδιακύμανσης, που συμβολίζεται S_{PW} συνδυάζοντας δυο μεθόδους, την κλασσική μέθοδο και τη μέθοδο παρατηρήσεων με ελλιπή δεδομένα.

Αρχικά για κάθε μεταβλητή $\boldsymbol{X}_j$ όπου $j = 1, 2, \ldots, p$ από την (1.12) υπολογίζεται η μέση τιμή της χρησιμοποιώντας όλες τις διαθέσιμες τιμές και από την (1.15) υπολογίζεται η διασπορά της $\boldsymbol{X}_j$. Οι διασπορές s_{jj}, που υπολογίζονται, αποτελούν τα διαγώνια στοιχεία του $p \times p$ πίνακα συνδιακύμανσης S_{PW}.

Στη συνέχεια, εφαρμόζουμε τη μέθοδο παρατηρήσεων με ελλιπή δεδομένα μεταξύ δυο μεταβλητών $\boldsymbol{X}_j$ και $\boldsymbol{X}_k$ με $j,k = 1, 2, \ldots, p$, $j \neq k$, προκειμένου το σύνολο να έχει όλες τις παρατηρήσεις πλήρεις και για τις δύο μεταβλητές του. Σε κάθε τέτοιο σύνολο (για κάθε ζεύγος μεταβλητών) υπολογίζουμε εκ νέου τη μέση τιμή της μεταβλητής από την (1.12) και εν συνεχεία τη συνδιακύμανση s_{jk} από την (1.17), $j,k = 1, 2, \ldots, p$, $j \neq k$, η οποία αποτελεί το μη διαγώνιο στοιχείο του πίνακα συνδιακύμανσης S_{PW} [26].

Σχόλιο 2.3

i. Το πλεονέκτημα της μεθόδου διαγραφής ανά ζεύγη είναι ότι αξιοποιεί όσο το δυνατών περισσότερα δεδομένα, καθώς αν και εξαιρεί μια παρατήρηση όταν έχει μια μη διαθέσιμη τιμή, χρησιμοποιεί τη συγκεκριμένη παρατήρηση κατά την ανάλυση των άλλων μεταβλητών.
ii. Το βασικό μειονέκτημα της μεθόδου είναι το γεγονός ότι δεν εγγυάται ότι ο πίνακας συνδιακύμανσης S_{PW} που προκύπτει, είναι θετικά ημιορισμένος,

καθώς κάθε εκτίμηση βασίζεται σε διαφορετικό υποσύνολο δεδομένων [13]. Έτσι, ο S_{PW} **δεν** είναι **πάντα** ένας **θετικά ημιορισμένος** πίνακας (βλέπε Παράδειγμα 2.6 στη συνέχεια).

iii. Επειδή κατά την εφαρμογή της μεθόδου διαγραφής ανά ζεύγη υπολογίστηκε ο πίνακας συνδιακύμανσης S_{PW}, σύμφωνα με το Πόρισμα 1.1 μπορεί να υπολογιστεί άμεσα και ο αντίστοιχος πίνακας συσχέτισης που θα συμβολίζεται R_{PW}. Σύμφωνα με το παραπάνω (ii) σχόλιο και την Πρόταση 1.12 ο πίνακας R_{LW} μπορεί να είναι **αόριστος**[18], γεγονός που εξαρτάται από τα πρόσημα των ιδιοτιμών του πίνακα S_{PW}. □

Παράδειγμα 2.5

Χρησιμοποιώντας το σύνολο με τα ελλιπή δεδομένα του Παραδείγματος 2.1 και εφαρμόζοντας τη μέθοδο διαγραφής ανά ζεύγη, να βρεθούν:

i. ο πίνακας συνδιακύμανσης S_{PW},

ii. ο πίνακας συσχέτισης R_{PW}, και

iii. να υπολογιστούν οι ιδιοτιμές του S_{PW}.

i. Αρχικά, υπολογίζουμε τις μέσες τιμές κάθε μεταβλητής $\boldsymbol{X}_1, \boldsymbol{X}_2$ και $\boldsymbol{X}_3$ με τη χρήση των διαθέσιμων παρατηρήσεων και της σχέσης (1.12):

$$\bar{\mathrm{x}}_1 = E(\boldsymbol{X}_1) = \frac{1+3+8+0+5+3+4+0}{8} = \frac{24}{8} = 3$$

$$\bar{\mathrm{x}}_2 = E(\boldsymbol{X}_2) = \frac{2+2+3+8+0+1+1}{7} = \frac{17}{7} = 2.428$$

$$\bar{\mathrm{x}}_3 = E(\boldsymbol{X}_3) = \frac{14+2+1+2+1+3+4+2}{8} = \frac{29}{8} = 3.625$$

Τα στοιχεία του πίνακα συνδιακύμανσης είναι:

$$s_{11} = \frac{1}{8-1}\sum_{i=1}^{8}(x_{i1} - \bar{\mathrm{x}}_1)^2 = \frac{52}{7} = 7.428$$

---

[18] Βλέπε Ορισμό 1.6.

$$s_{22} = \frac{1}{7-1}\sum_{i=1}^{7}(x_{i2} - \overline{\mathrm{x}}_2)^2 = \frac{41.714}{6} = 6.952$$

$$s_{33} = \frac{1}{8-1}\sum_{i=1}^{8}(x_{i3} - \overline{\mathrm{x}}_3)^2 = \frac{129.874}{7} = 18.553$$

Σύμφωνα με τη μέθοδο διαγραφής ανά ζεύγη, για τον υπολογισμό των μη διαγωνίων στοιχείων του πίνακα συνδιακύμανσης θα χρησιμοποιηθούν, για κάθε ζεύγος μεταβλητών, όλες οι περιπτώσεις με πλήρεις παρατηρήσεις και στις δυο μεταβλητές, το οποίο δίνει τα ακόλουθα σύνολα δεδομένων.

α)

$\boldsymbol{X}_1$	$\boldsymbol{X}_2$
1	2
8	2
0	3
3	0
4	1
0	1

β)

$\boldsymbol{X}_1$	$\boldsymbol{X}_3$
1	14
3	2
8	2
3	3
4	4
0	2

γ)

$\boldsymbol{X}_2$	$\boldsymbol{X}_3$
2	14
2	2
8	1
0	3
1	4
1	2

Σε καθένα από τα παραπάνω σύνολα δεδομένων, για κάθε μεταβλητή υπολογίζεται η μέση τιμή της και εν συνεχεία η αντίστοιχη συνδιακύμανση αυτών των μεταβλητών.

Για το **(α)**:

$$\overline{\mathrm{x}}_1 = E(\boldsymbol{X}_1\ ) = \frac{1+8+0+3+4+0}{6} = \frac{16}{6} = 2.666$$

$$\overline{\mathrm{x}}_2 = E(\boldsymbol{X}_2\ ) = \frac{2+2+3+0+1+1}{6} = \frac{9}{6} = 1.5$$

$$s_{12} = \frac{1}{6-1}\sum_{i=1}^{6}(x_{i1} - \overline{\mathrm{x}}_1)(x_{i2} - \overline{\mathrm{x}}_2) = \frac{-2}{5} = -0.4$$

Για το **(β)**:

$$\bar{\mathrm{x}}_1 = E(\boldsymbol{X}_1\ ) = \frac{1+3+8+3+4+0}{6} = \frac{19}{6} = 3.1666$$

$$\bar{\mathrm{x}}_3 = E(\boldsymbol{X}_3) = \frac{14+2+2+3+4+2}{6} = \frac{27}{6} = 4.5$$

$$s_{13} = \frac{1}{6-1}\sum_{i=1}^{6}(x_{i1}-\bar{\mathrm{x}}_1)(x_{i3}-\bar{\mathrm{x}}_3) = \frac{-24.5}{5} = -4.9$$

Για το **(γ)**:

$$\bar{\mathrm{x}}_2 = E(\boldsymbol{X}_2\ ) = \frac{2+2+8+0+1+1}{6} = \frac{14}{6} = 2.333$$

$$\bar{\mathrm{x}}_3 = E(\boldsymbol{X}_3) = \frac{14+2+1+3+4+2}{6} = \frac{26}{6} = 4.333$$

$$s_{23} = \frac{1}{6-1}\sum_{i=1}^{6}(x_{i2}-\bar{\mathrm{x}}_2)(x_{i3}-\bar{\mathrm{x}}_3) = \frac{-14.666}{5} = -2.933$$

Άρα, ο πίνακας συνδιακύμανσης S_{PW} είναι ο:

$$S_{PW} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix} = \begin{bmatrix} 7.428 & -0.4 & -4.9 \\ -0.4 & 6.952 & -2.933 \\ -4.9 & -2.933 & 18.553 \end{bmatrix}$$

ii. Σύμφωνα με το Πόρισμα 1.1 και τα στοιχεία του παραπάνω πίνακα συνδιακύμανσης S_{PW}, ο πίνακας συσχέτισης R_{PW} που προκύπτει είναι:

$$R_{LW} = (V^{1/2})^{-1} S_{LW} (V^{1/2})^{-1}$$

$$= \begin{bmatrix} \sqrt{7.428} & 0 & 0 \\ 0 & \sqrt{6.952} & 0 \\ 0 & 0 & \sqrt{18.553} \end{bmatrix}^{-1} \begin{bmatrix} 7.428 & -0.4 & -4.9 \\ -0.4 & 6.952 & -2.933 \\ -4.9 & -2.933 & 18.553 \end{bmatrix} \begin{bmatrix} \sqrt{7.428} & 0 & 0 \\ 0 & \sqrt{6.952} & 0 \\ 0 & 0 & \sqrt{18.553} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1 & -0.0557 & -0.4174 \\ -0.0557 & 1 & -0.2583 \\ -0.4174 & -0.2583 & 1 \end{bmatrix}$$

iii. Οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{PW} υπολογίζονται από την (1.3) και έχουμε:

$$\det(S_{PW} - \lambda I) = 0 \Rightarrow -\lambda^3 + 32.933\lambda^2 - 285.659\lambda + 712.784 = 0$$

Επομένως, οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{PW} είναι:

$$\lambda_1 = 4.563,\ \lambda_2 = 7.476,\ \lambda_3 = 20.894.$$

Επειδή οι ιδιοτιμές του πίνακα είναι θετικές, ο συμμετρικός πίνακας S_{PW} είναι θετικά ημιορισμένος (ιδιότητα (viii), Πρόταση 1.3). ◊

Παράδειγμα 2.6

Χρησιμοποιώντας το σύνολο με τα ελλιπή δεδομένα του Παραδείγματος 2.2 και εφαρμόζοντας τη μέθοδο διαγραφής ανά ζεύγη, να βρεθούν:

i. ο πίνακας συνδιακύμανσης S_{PW},
ii. ο πίνακας συσχέτισης R_{PW}, και
iii. να υπολογιστούν οι ιδιοτιμές των S_{PW} και R_{PW}.

i. Αρχικά, υπολογίζουμε τις μέσες τιμές κάθε μεταβλητής $\boldsymbol{X}_1, \boldsymbol{X}_2$ και $\boldsymbol{X}_3$ με τη χρήση των διαθέσιμων παρατηρήσεων και την (1.12):

$$\overline{\mathrm{x}}_1 = E(\boldsymbol{X}_1\ ) = \frac{1+3+8+0+5+2+3+4+0}{9} = \frac{26}{9} = 2.888$$

$$\overline{\mathrm{x}}_2 = E(\boldsymbol{X}_2\ ) = \frac{1+2+7+1}{4} = \frac{11}{4} = 2.75$$

$$\overline{\mathrm{x}}_3 = E(\boldsymbol{X}_3) = \frac{2+1+3+8+1+3+4+2}{8} = \frac{24}{8} = 3$$

Τα στοιχεία του πίνακα συνδιακύμανσης S_{PW} είναι:

$$s_{11} = \frac{1}{9-1}\sum_{i=1}^{9}(x_{i1} - \overline{x}_1)^2 = \frac{52.888}{8} = 6.611$$

$$s_{22} = \frac{1}{4-1}\sum_{i=1}^{4}(x_{i2} - \overline{x}_2)^2 = \frac{24.75}{3} = 8.25$$

$$s_{33} = \frac{1}{8-1}\sum_{i=1}^{8}(x_{i3} - \overline{x}_3)^2 = \frac{36}{7} = 5.143$$

Σύμφωνα με τη μέθοδο διαγραφής ανά ζεύγη, για τον υπολογισμό των μη διαγωνίων στοιχείων του πίνακα συνδιακύμανσης θα χρησιμοποιηθούν, για κάθε ζεύγος μεταβλητών, όλες οι περιπτώσεις με πλήρεις παρατηρήσεις και στις δυο μεταβλητές, το οποίο δίνει τα ακόλουθα σύνολα δεδομένων.

α)

$\boldsymbol{X}_1$	$\boldsymbol{X}_2$
3	1
8	2
5	7
4	1

β)

$\boldsymbol{X}_1$	$\boldsymbol{X}_3$
3	2
0	3
5	8
2	1
3	3
4	4
0	2

γ)

$\boldsymbol{X}_2$	$\boldsymbol{X}_3$
1	2
7	8
1	4

Σε καθένα από τα παραπάνω σύνολα δεδομένων, για κάθε μεταβλητή υπολογίζεται η μέση τιμή της και εν συνεχεία η αντίστοιχη συνδιακύμανση αυτών των μεταβλητών.

Για το **(α)**:

$$\overline{\mathrm{x}}_1 = E(\boldsymbol{X}_1\ ) = \frac{3+8+5+4}{4} = \frac{20}{4} = 5$$

$$\overline{\mathrm{x}}_2 = E(\boldsymbol{X}_2\ ) = \frac{1+2+7+1}{4} = \frac{11}{4} = 2.75$$

$$s_{12} = \frac{1}{4-1}\sum_{i=1}^{4}(x_{i1}-\overline{x}_1)(x_{i2}-\overline{x}_2) = \frac{4}{3} = 1.333$$

Για το **(β)**:

$$\overline{\mathrm{x}}_1 = E(\boldsymbol{X}_1\ ) = \frac{3+0+5+2+3+4+0}{7} = \frac{17}{7} = 2.428$$

$$\overline{\mathrm{x}}_3 = E(\boldsymbol{X}_3) = \frac{2+3+8+1+3+4+2}{7} = \frac{23}{7} = 3.285$$

$$s_{13} = \frac{1}{7-1}\sum_{i=1}^{7}(x_{i1}-\overline{x}_1)(x_{i3}-\overline{x}_3) = \frac{17.143}{6} = 2.857$$

Για το **(γ)**:

$$\overline{\mathrm{x}}_2 = E(\boldsymbol{X}_2\ ) = \frac{1+7+1}{3} = \frac{9}{3} = 3$$

$$\overline{\mathrm{x}}_3 = E(\boldsymbol{X}_3) = \frac{2+8+4}{3} = \frac{14}{3} = 4.666$$

$$s_{23} = \frac{1}{3-1}\sum_{i=1}^{3}(x_{i2}-\overline{x}_2)(x_{i3}-\overline{x}_3) = \frac{20.44}{2} = 10.22$$

Άρα, ο πίνακας συνδιακύμανσης S_{PW} είναι:

$$S_{PW} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix} = \begin{bmatrix} 6.611 & 1.333 & 2.857 \\ 1.333 & 8.25 & 10.22 \\ 2.857 & 10.22 & 5.143 \end{bmatrix}$$

ii. Σύμφωνα με το Πόρισμα 1.1 και τα στοιχεία του παραπάνω πίνακα συνδιακύμανσης S_{PW}, ο πίνακας συσχέτισης R_{PW} που προκύπτει είναι:

$$R_{LW} = (V^{1/2})^{-1} S_{LW} (V^{1/2})^{-1}$$

$$= \begin{bmatrix} \sqrt{6.611} & 0 & 0 \\ 0 & \sqrt{8.25} & 0 \\ 0 & 0 & \sqrt{5.143} \end{bmatrix}^{-1} \begin{bmatrix} 6.611 & 1.333 & 2.857 \\ 1.333 & 8.25 & 10.22 \\ 2.857 & 10.22 & 5.143 \end{bmatrix} \begin{bmatrix} \sqrt{6.611} & 0 & 0 \\ 0 & \sqrt{8.25} & 0 \\ 0 & 0 & \sqrt{5.143} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1 & 0.1805 & 0.49 \\ 0.1805 & 1 & 1.569 \\ 0.49 & 1.569 & 1 \end{bmatrix}$$

iii. Οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{PW} υπολογίζονται από την (1.3) και έχουμε:

$$\det(S_{PW} - \lambda I) = 0 \Rightarrow -\lambda^3 + 20.004\lambda^2 - 16.583\lambda - 408.641 = 0$$

Επομένως, οι ιδιοτιμές του πίνακα συνδιακύμανσης είναι οι:

$$\lambda_1 = -3.809,\ \lambda_2 = 6.034,\ \lambda_3 = 17.778$$

Παρατηρούμε ότι ο πίνακας συνδιακύμανσης S_{PW} που προέκυψε είναι αόριστος[19] καθώς έχει θετικές και αρνητικές ιδιοτιμές, επαληθεύοντας το Σχόλιο 2.3 (ii).

---

[19] Βλέπε Ορισμό 1.6.

Οι ιδιοτιμές του πίνακα συσχέτισης R_{PW} υπολογίζονται από τη χαρακτηριστική εξίσωση:

$$\det(R_{PW} - \lambda I) = 0 \Rightarrow -\lambda^3 + 3\lambda^2 - 0.2657\lambda - 1.4568 = 0$$

Επομένως, οι ιδιοτιμές του πίνακα συσχέτισης R_{PW} είναι οι:

$$\lambda_1 = -0.6003,\ \lambda_2 = 0.8981,\ \lambda_3 = 2.7022$$

Παρατηρούμε ότι ο πίνακας συσχέτισης R_{PW} που προέκυψε είναι αόριστος καθώς έχει θετικές και αρνητικές ιδιοτιμές, επαληθεύοντας το Σχόλιο 2.3 (iii). ◊

Στα ακόλουθα παραδείγματα παρουσιάζονται τα αποτελέσματα που δίνει η υλοποίηση σε Stata του κώδικα, (βλέπε Παράρτημα Α), χρησιμοποιώντας τις εντολές του πακέτου για να υπολογιστεί ο πίνακας συνδιακύμανσης με σύνολα δεδομένων, όπως δίνονται στο Παράρτημα Γ.

Παράδειγμα 2.7

Έστω το σύνολο δεδομένων του Παραρτήματος Γ.1. Χρησιμοποιώντας κατάλληλα τμήματα του κώδικα από το Παράρτημα Α να υπολογιστούν:

i. ο πίνακας συνδιακύμανσης S των δεδομένων,

ii. ο πίνακας συνδιακύμανσης S_{PW} και οι ιδιοτιμές του.

i. Από το τμήμα του κώδικα στο Παράρτημα Α

```
correlate var1-var3,cov
matrix Cf = r(C)
mat list Cf
```

(Υπολογισμός του πίνακα συνδιακύμανσης σύμφωνα με τον Ορισμό 1.18)

προκύπτει ο πίνακας συνδιακύμανσης S για το σύνολο δεδομένων:

$$S = C_f = \begin{bmatrix} 380.73 & 145.26 & 385.61 \\ 145.26 & 101.29 & 208.45 \\ 385.61 & 208.45 & 1540.9 \end{bmatrix}$$

ii. Από το τμήμα του κώδικα στο Παράρτημα Α

```
forvalues x=1/3 {
   forvalues y=1/3{
        correlate var`x' var`y',cov
```

(Δημιουργία πίνακα συνδιακύμανσης S_{PW} με μέθοδο διαγραφής ανά ζεύγη)

```
            local var`x'`y'= r(cov_12)
        }
    }

    matrix Co =(`var11',`var12', `var13' \ `var21',`var22',
   `var23' \ `var31',`var32' ,`var33' )
    mat list Co
```

προκύπτει ο πίνακας συνδιακύμανσης S_{PW} (μέθοδος διαγραφής ανά ζεύγη):

$$S_{PW} = C_0 = \begin{bmatrix} 468.162 & 231.002 & 475.859 \\ 231.002 & 110.032 & 161.064 \\ 475.859 & 161.064 & 1457.54 \end{bmatrix}$$

Τέλος, από το τμήμα του κώδικα στο Παράρτημα Α έχουμε

```
matrix symeigen X v = Co
 mat list v
```

(Υπολογισμός των ιδιοδιανυσμάτων και ιδιοτιμών του πίνακα συνδιακύμανσης S_{PW})

προκύπτουν οι ιδιοτιμές του πίνακα συνδιακύμανσης S_{PW} :

$$\lambda_1 = -7.20,\ \lambda_2 = 358.01,\ \lambda_3 = 1684.92$$

Επειδή οι ιδιοτιμές είναι άλλες θετικές και άλλες αρνητικές, ο πίνακας συνδιακύμανσης S_{PW} είναι αόριστος[20], γεγονός που επαληθεύει το Σχόλιο 2.3 (ii). ◊

Σχόλιο 2.4

Συνοψίζοντας, παρατηρούμε ότι η μέθοδος παρατηρήσεων με ελλιπή δεδομένα καθώς και η μέθοδος ΕΜ υπολογίζουν ενός θετικά ημιορισμένους πίνακες συνδιακύμανσης και συσχέτισης (βλέπε Σχόλιο 2.1 (iii) και 2.2 (iii) καθώς και τα Παραδείγματα 2.1-2.4). Όπως αναφέρθηκε στο Σχόλιο 2.3 (ii, iv) και επιβεβαιώθηκε με τα Παραδείγματα 2.5-2.7, η μέθοδος διαγραφής ανά ζεύγη δε δίνει πάντα θετικά ημιορισμένους πίνακες συνδιακύμανσης και συσχέτισης, γεγονός που δεν επιτρέπει περαιτέρω στατιστική ανάλυση των πολυμεταβλητών δεδομένων. Το πρόβλημα αυτό θα μας απασχολήσει στο επόμενο κεφάλαιο όπου θα προταθούν μέθοδοι διόρθωσης των αρνητικών ιδιοτιμών των στατιστικών πινάκων. □

---

[20] Βλέπε Ορισμό 1.6.

$3^{ο}$ ΚΕΦΑΛΑΙΟ

ΜΕΘΟΔΟΙ ΔΙΑΤΑΡΑΧΗΣ ΙΔΙΟΤΙΜΩΝ ΣΥΜΜΕΤΡΙΚΩΝ ΠΙΝΑΚΩΝ

Σύμφωνα με το Σχόλιο 2.4 η ιδιότητα[21] των πινάκων συνδιακύμανσης και συσχέτισης να είναι πάντα θετικά ημιορισμένοι κάποιες φορές δεν ικανοποιείται. Το γεγονός αυτό, σύμφωνα με τη βιβλιογραφία [17], συνιστά σοβαρό πρόβλημα σε διάφορες πολυμεταβλητές τεχνικές ανάλυσης δεδομένων, όπως για παράδειγμα, στην παλινδρόμηση (regression) και στην ανάλυση κατά παράγοντες (factor analysis).

Στο κεφάλαιο αυτό παρατίθενται μέθοδοι διαταραχής ιδιοτιμών οι οποίες τιτλοφορούνται ως ακολούθως

1. μερικής φασματικής ανάλυσης,
2. μεταφοράς, και
3. μέσω πίνακα συσχέτισης

προκειμένου να παρακαμφθεί αοριστία των παραπάνω πινάκων και να προκύψουν θετικά ημιορισμένοι πίνακες.

Στο τέλος του κεφαλαίου συγκρίνονται οι παραπάνω μέθοδοι διαταραχής βάσει:

1. νόρμας Frobenius,
2. κριτηρίου διαφοράς με χρήση απόλυτων τιμών, και
3. γραμμικής παλινδρόμησης.

Η σύγκριση των μεθόδων θα συμβάλλει στην εξαγωγή συμπερασμάτων αναφορικά με το ποια μέθοδος είναι βέλτιστη.

Αξίζει να σημειωθεί ότι τα σύνολα δεδομένων που χρησιμοποιήθηκαν για την στατιστική ανάλυση παρατίθενται στο Παράρτημα Γ. Οι κώδικες που αναπτύχθηκαν για Stata προκειμένου να γίνουν οι υπολογισμοί των πινάκων σε κάθε μέθοδο διαταραχής των ιδιοτιμών καθώς και οι συγκρίσεις τους παρατίθενται στα Παραρτήματα Α και Β.

---

[21] Βλέπε ιδιότητα (i) Πρόταση 1.7.

3.1. Μέθοδοι διαταραχής ιδιοτιμών

3.1.1 Μέθοδος μερικής φασματικής ανάλυσης

Όπως αναφέρθηκε στο Κεφάλαιο 1, η φασματική ανάλυση ενός συμμετρικού πίνακα υπολογίζεται με τη χρήση όλων των ιδιοτιμών $\lambda_1, \lambda_2, \ldots, \lambda_p$, που προκύπτουν ως ρίζες του χαρακτηριστικού πολυωνύμου του πίνακα, και των αντίστοιχων ορθογώνιων ιδιοδιανυσμάτων τους $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$, (Θεώρημα 1.2). Επειδή οι ιδιοτιμές του συμμετρικού πίνακα είναι πραγματικοί αριθμοί (βλέπε ιδιότητα (vii), Πρόταση 1.1) μπορούμε να θεωρήσουμε ότι έχουν την ακόλουθη διάταξη:

$$\lambda_1 \le \lambda_2 \le \ldots \le 0 \le \ldots \le \lambda_p . \tag{3.1}$$

Ο όρος «μερική» φασματική ανάλυση οφείλεται στο γεγονός ότι δε γίνεται χρήση όλων των ιδιοτιμών του συμμετρικού πίνακα αλλά μόνο των μη-μηδενικών.

Η μέθοδος προτείνει την αντικατάσταση των αρνητικών ιδιοτιμών με μηδέν. Έτσι, υπολογίζεται ένας νέος θετικά ημιορισμένος πίνακας, ο οποίος έχει το ίδιο πλήθος θετικών ιδιοτιμών με τον αρχικό και το μηδέν είναι ιδιοτιμή του με αλγεβρική πολλαπλότητα ίση με το πλήθος των αρχικών μηδενικών και των αρνητικών ιδιοτιμών του αρχικού πίνακα.

Αν σε ένα σύνολο p-πολυμεταβλητών δεδομένων ($\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ οι μεταβλητές του), χρησιμοποιώντας τη μέθοδο διαγραφής ανά ζεύγη, προκύψει πίνακας συνδιακύμανσης S_{PW} αόριστος[22], διαταράσσουμε τις αρνητικές του ιδιοτιμές, εφαρμόζοντας την ακόλουθη πρόταση. Η απόδειξη της οποίας βασίζεται στο Θεώρημα 1.2 και την ιδιότητα (viii) της Πρότασης 1.3.

Πρόταση 3.1

Έστω S_{PW} πίνακας συνδιακύμανσης με ιδιοτιμές όπως (3.1) και $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$ τα αντίστοιχα ορθογώνια ιδιοδιανύσματά τους τότε, ο πίνακας S_{eig}

$$S_{eig} = \sum_{\lambda_i \ge 0} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T \tag{3.2}$$

είναι ένας θετικά ημιορισμένος πίνακας.

---

[22] Δηλαδή έχει και αρνητικές ιδιοτιμές (βλέπε Ορισμό 1.6 και Πρόταση 1.3).

Παράδειγμα 3.1

Έστω

$$S_{PW}=\begin{bmatrix} 6.611 & 1.333 & 2.857 \\ 1.333 & 8.25 & 10.22 \\ 2.857 & 10.22 & 5.143 \end{bmatrix}$$

ο αόριστος πίνακας συνδιακύμανσης του Παραδείγματος 2.6. Να υπολογιστεί θετικά ημιορισμένος πίνακας με τη μέθοδο μερικής φασματικής ανάλυσης.

Επειδή οι ιδιοτιμές του πίνακα S_{PW} είναι

$$\lambda_1=-3.809\,,\ \lambda_2=6.034\,,\ \lambda_3=17.778$$

και τα αντίστοιχα ορθογώνια ιδιοδιανύσματα τους είναι στήλες του ακόλουθου πίνακα

$$U=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{u}_3\end{bmatrix}=\begin{bmatrix} -0.1284 & -0.9595 & -0.2508 \\ -0.6329 & 0.274 & -0.7241 \\ 0.7635 & 0.0658 & -0.6424 \end{bmatrix},$$

σύμφωνα με την Πρόταση 3.1 και την (3.2), η ιδιοτιμή λ_1 θα αντικατασταθεί με μηδέν και ο νέος πίνακας είναι:

$$S_{eig}=\sum_{\lambda_i\geq 0}\lambda_i\boldsymbol{u}_i\boldsymbol{u}_i^T=\lambda_2\boldsymbol{u}_2\boldsymbol{u}_2^T+\lambda_3\boldsymbol{u}_3\boldsymbol{u}_3^T=\begin{bmatrix} 6.6738 & 1.6425 & 2.4836 \\ 1.6425 & 9.7757 & 8.3794 \\ 2.4836 & 8.3794 & 7.3635 \end{bmatrix}$$

Προφανώς, ο πίνακας S_{eig} είναι θετικά ημιορισμένος από την παραπάνω φασματική ανάλυση. ◊

3.1.2 Μέθοδος μεταφοράς

Μια άλλη μέθοδος που μπορεί να χρησιμοποιηθεί για τη μετατροπή αόριστων πινάκων σε θετικά ημιορισμένους, είναι η μέθοδος μεταφοράς. Η μέθοδος προτείνει τη «μεταφορά» της μικρότερης ιδιοτιμής λίγο μετά το μηδέν [1], θεωρώντας ότι οι ιδιοτιμές του συμμετρικού πίνακα έχουν τη διάταξη όπως στην (3.1), δηλαδή,

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq 0 \leq \ldots \leq \lambda_p .$$

Η ιδέα της μεθόδου απορρέει από την ιδιότητα της Πρότασης 1.2 για τον πίνακα αυτόν

$$S_{met} = S_{PW} + (\varepsilon - \lambda_1) I_n \tag{3.3}$$

όπου S_{PW} ένας αόριστος πίνακας συνδιακύμανσης, που προκύπτει από την εφαρμογή της μεθόδου διαγραφής ανά ζεύγη, I_n ο μοναδιαίος πίνακας, ε μια μικρή θετική σταθερή τιμή (για παράδειγμα $\varepsilon = 10^{-6}$) και λ_1 η μικρότερη αρνητική ιδιοτιμή του πίνακα S_{PW}.

Ο νέος πίνακας S_{met}, που προκύπτει κατά τη «μεταφορά» του αόριστου πίνακα συνδιακύμανσης, είναι θετικά ημιορισμένος πίνακας, επειδή οι ιδιοτιμές του είναι

$$\lambda_{met} = \lambda_i + \varepsilon - \lambda_1 \, , \; i = 1, 2, \ldots, p \tag{3.4}$$

με $\lambda_i \in \sigma(S_{PW})$.

Παράδειγμα 3.2

Έστω

$$S_{PW} = \begin{bmatrix} 6.611 & 1.333 & 2.857 \\ 1.333 & 8.25 & 10.22 \\ 2.857 & 10.22 & 5.143 \end{bmatrix}$$

ο αόριστος πίνακας συνδιακύμανσης του Παραδείγματος 2.6. Να υπολογιστεί θετικά ημιορισμένος πίνακας με τη μέθοδο μεταφοράς, όπου $\varepsilon = 10^{-6}$.

Σύμφωνα με την (3.3) ο πίνακας S_{met} είναι:

$$S_{met} = S + (\varepsilon - \lambda) I = S_{PW} + (10^{-6} - (-3.809)) I_n = \begin{bmatrix} 10.4199 & 1.333 & 2.857 \\ 1.333 & 12.0589 & 10.22 \\ 2.857 & 10.22 & 8.9519 \end{bmatrix}$$

Επειδή οι ιδιοτιμές του πίνακα S_{PW} είναι

$$\lambda_1 = -3.809\,,\ \lambda_2 = 6.034\,,\ \lambda_3 = 17.778$$

σύμφωνα με την (3.4) οι ιδιοτιμές του S_{met} είναι

$$\lambda_{met1} = 0.000001, \lambda_{met2} = 9.8435 \text{ και } \lambda_{met3} = 21.5872$$

Προφανώς, ο πίνακας S_{met} είναι θετικά ημιορισμένος.

Επίσης, παρατηρούμε ότι ο πίνακας S_{met} που προέκυψε από τη μέθοδο της μεταφοράς διαφέρει από τον αρχικό πίνακα S_{PW} μόνο στα διαγώνια στοιχεία. Το συμπέρασμα, λοιπόν, που απορρέει είναι ότι η μέθοδος της μεταφοράς διαταράσσει μόνο τα στοιχεία της διαγωνίου και τις ιδιοτιμές. ◊

3.1.3 Μέθοδος μέσω πίνακα συσχέτισης

Όπως αναφέρθηκε στα Σχόλια 2.3 και 2.4 κατά τη στατιστική ανάλυση με τη μέθοδο διαγραφής ανά ζεύγη μπορεί να προκύψει πίνακας συσχέτισης R_{PW} αόριστος, Έστω $r_1, r_2, \ldots, r_p \in \sigma(R_{PW})$ να είναι οι ιδιοτιμές του πίνακα R_{PW} με

$$r_1 \leq r_2 \leq \ldots \leq 0 \leq \ldots \leq r_p \tag{3.5}$$

και $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_p$ τα αντίστοιχα ορθογώνια ιδιοδιανύσματα, τα οποία αποτελούν τις στήλες ενός πίνακα W.

Η βασική ιδέα της μεθόδου είναι να επεκτείνει τις μεθόδους που αναπτύχθηκαν για έναν αόριστο πίνακα συνδιακύμανσης S_{PW}, στις Ενότητες 3.1.1 και 3.1.2, για έναν αόριστο πίνακα συσχέτισης R_{PW}. Έτσι, στις [23, 27] προτείνεται η αντικατάσταση όλων των αρνητικών ιδιοτιμών του R_{PW} με την τιμή μηδέν και στο πλαίσιο της παρούσας πτυχιακής προτείνεται η αντικατάσταση όλων των αρνητικών ιδιοτιμών του R_{PW} από ένα μικρό θετικό αριθμό. Εναλλακτικά, οι αρνητικές ιδιοτιμές μπορούν να αντικατασταθούν από τις απόλυτες τιμές τους.

Ο ακόλουθος αλγόριθμος είναι αυτός που αναπτύχθηκε στην παρούσα πτυχιακή.

Αλγόριθμος 3.1

1. Εύρεση ιδιοτιμών r_i του πίνακα συσχέτισης R_{PW} όπως στην (3.5) και αντίστοιχων ιδιοδιανυσμάτων $\boldsymbol{w}_i$ και τοποθέτηση τους ως στήλες σε έναν πίνακα W.
2. Αντικατάσταση όλων των αρνητικών ιδιοτιμών με $\varepsilon > 0$ μικρό.
3. Κατασκευή R' τέτοιου ώστε

   $$R' = WDW^T$$

   όπου $D = diag(\varepsilon, \varepsilon, \ldots, r_p)$.
4. Υπολογισμός νέου πίνακα συσχέτισης.

   i. $t_{ii} = \dfrac{1}{\sqrt{r'_{ii}}}$, όπου r'_{ii} διαγώνια στοιχεία του πίνακα R'

για κάθε $i = 1, 2, \ldots, p$, και θέσε $\boldsymbol{t} = [t_{11} \ t_{22} \ \ldots \ t_{pp}]^T$.

ii. $T_{new} = \boldsymbol{TT}^T$

iii. $R_{new} = R' \circ T_{new}$

όπου $\circ$ σημειώνεται το γινόμενο Hadamard[23].

Παράδειγμα 3.3

Έστω

$$R_{PW} = \begin{bmatrix} 1 & 0.1805 & 0.49 \\ 0.1805 & 1 & 1.569 \\ 0.49 & 1.569 & 1 \end{bmatrix}$$

ο αόριστος πίνακας συσχέτισης του Παραδείγματος 2.6. Χρησιμοποιώντας τον Αλγόριθμο 3.1 να υπολογιστεί θετικά ορισμένος πίνακας συσχέτισης για $\varepsilon = 10^{-6}$.

1. Οι ιδιοτιμές του πίνακα R_{PW} είναι οι:

   $$r_1 = -0.6003,\ r_2 = 0.8981,\ r_3 = 2.7022$$

   ενώ τα αντίστοιχα ιδιοδιανύσματα είναι τα ακόλουθα:

   $$W = \begin{bmatrix} 0.1416 & 0.9523 & 0.2704 \\ 0.6848 & -0.2915 & 0.6678 \\ -0.7148 & -0.0906 & 0.6934 \end{bmatrix}$$

2. Αντικαθιστούμε τις αρνητικές ιδιοτιμές με το $\varepsilon = 10^{-6}$. Ο διαγώνιος πίνακας με μη αρνητικές ιδιοτιμές είναι:

   $$D = \begin{bmatrix} 10^{-6} & 0 & 0 \\ 0 & 0.8981 & 0 \\ 0 & 0 & 2.7022 \end{bmatrix}$$

---

[23] Βλέπε Ορισμό 1.9.

3. Κατασκευάζουμε τον πίνακα συσχέτισης:

$$R' = WDW^T$$

$$= \begin{bmatrix} 0.16 & 0.95 & 0.26 \\ 0.68 & -0.29 & 0.67 \\ -0.71 & -0.07 & 0.69 \end{bmatrix} \begin{bmatrix} 10^{-6} & 0 & 0 \\ 0 & 0.8981 & 0 \\ 0 & 0 & 2.7022 \end{bmatrix} \begin{bmatrix} 0.16 & 0.95 & 0.26 \\ 0.68 & -0.29 & 0.67 \\ -0.71 & -0.07 & 0.69 \end{bmatrix}^T$$

$$= \begin{bmatrix} 1.012 & 0.2387 & 0.4292 \\ 0.2387 & 1.2815 & 1.2751 \\ 0.4292 & 1.2751 & 1.3067 \end{bmatrix}$$

Παρατηρούμε ότι ο νέος πίνακας συσχέτισης R' δεν πληροί τη βασική ιδιότητα[24] των πινάκων συσχέτισης, σύμφωνα με την οποία τα διαγώνια στοιχεία του πίνακα θα έπρεπε να είναι ίσα με τη μονάδα.

4. Υπολογισμός νέου πίνακα συσχέτισης.

i. $\boldsymbol{t} = [t_{11} \ \ t_{22} \ \ t_{33}]^T = \begin{bmatrix} 0.994 \\ 0.8833 \\ 0.8748 \end{bmatrix}$

ii. $T_{new} = \boldsymbol{t}\,\boldsymbol{t}^T = \begin{bmatrix} 0.9881 & 0.8781 & 0.8696 \\ 0.8781 & 0.7803 & 0.7728 \\ 0.8696 & 0.7728 & 0.7653 \end{bmatrix}$

iii. $R_{new} = R' \circ T_{new} = \begin{bmatrix} 1 & 0.21 & 0.373 \\ 0.21 & 1 & 0.985 \\ 0.373 & 0.985 & 1 \end{bmatrix}$

Οι ιδιοτιμές του νέου πίνακα συσχέτισης R_{new} είναι οι ακόλουθες:

$$r_{new1} = 0.0005, \quad r_{new2} = 0.864 \quad \text{και} \quad r_{new3} = 2.1355.$$

Άρα, ο νέος πίνακας συσχέτισης είναι R_{new} θετικά ορισμένος καθώς $r_{newi} > 0$. ◊

---

[24] Βλέπε ιδιότητα (iii) της Πρότασης 1.8.

Σχόλιο 3.1

Αξιοποιώντας τον αόριστο πίνακα συνδιακύμανσης S_{PW} καθώς και τον αντίστοιχο πίνακα συσχέτισης R_{PW} (βλέπε Πόρισμα 1.1 ή Ορισμό 1.20) και εφαρμόζοντας τη μέθοδο μέσω πίνακα συσχέτισης κατασκευάζεται πίνακας R_{new}. Χρησιμοποιώντας τον R_{new}, τις διασπορές από τον S_{PW}, και τον τύπο (1.23) της Πρότασης 1.10, υπολογίζεται πίνακας που στη συνέχεια συμβολίζεται S_R. □

Παράδειγμα 3.4

Έστω ο πίνακας συνδιακύμανσης S_{PW} του Παραδείγματος 2.6

$$S_{PW} = \begin{bmatrix} 6.611 & 1.333 & 2.857 \\ 1.333 & 8.25 & 10.22 \\ 2.857 & 10.22 & 5.143 \end{bmatrix}$$

και ο πίνακας R_{new} που προέκυψε από τη μέθοδο μέσω πίνακα συσχέτισης του Παραδείγματος 3.3

$$R_{new} = \begin{bmatrix} 1 & 0.21 & 0.373 \\ 0.21 & 1 & 0.985 \\ 0.373 & 0.985 & 1 \end{bmatrix}$$

Σύμφωνα με το Σχόλιο 3.1 ο πίνακας S_R είναι:

$$\begin{aligned} S_R &= V^{\frac{1}{2}} R_{new} V^{\frac{1}{2}} \\ &= \begin{bmatrix} \sqrt{6.611} & 0 & 0 \\ 0 & \sqrt{8.25} & 0 \\ 0 & 0 & \sqrt{5.143} \end{bmatrix} \begin{bmatrix} 1 & 0.21 & 0.373 \\ 0.21 & 1 & 0.985 \\ 0.373 & 0.985 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{6.611} & 0 & 0 \\ 0 & \sqrt{8.25} & 0 \\ 0 & 0 & \sqrt{5.143} \end{bmatrix}^T \\ &= \begin{bmatrix} 6.611 & 1.5509 & 2.175 \\ 1.5509 & 8.2500 & 6.4161 \\ 2.175 & 6.4161 & 5.143 \end{bmatrix} \end{aligned}$$

Οι ιδιοτιμές του πίνακα S_R είναι οι:

$\lambda_1 = 0.0029$, $\lambda_2 = 5.8292$ και $\lambda_3 = 14.1719$ (δηλ. $\lambda_i \geq 0$). ◊

Παράδειγμα 3.5

Έστω το σύνολο δεδομένων του Παραρτήματος Γ.1 και ο αόριστος πίνακας συνδιακύμανσης $S_{PW} = C_o$ του Παραδείγματος 2.7. Χρησιμοποιώντας κατάλληλα τμήματα του κώδικα από το Παράρτημα Β να υπολογιστούν οι πίνακες συνδιακύμανσης που προέρχονται από τις ακόλουθες μεθόδους:

i. μερικής φασματικής ανάλυσης
ii. μεταφοράς και
iii. μεθόδου μέσω πίνακα συσχέτισης.

i.

```
matrix Cnew1=v[1,1]*X[1..3,1]*X[1..3,1]'+
v[1,2]*X[1..3,2]*X[1..3,2]'
```

(Υλοποίηση μεθόδου μερικής φασματικής ανάλυσης)

Ο πίνακας S_{eig} που προκύπτει από τη χρήση της μεθόδου της μερικής φασματικής ανάλυσης είναι:

$$S_{eig} = C_{new1} = \begin{bmatrix} 469.86 & 227.95 & 475.64 \\ 227.95 & 115.51 & 161.45 \\ 475.64 & 161.45 & 1457.6 \end{bmatrix}$$

```
scalar l=abs(v[1,3])+0.0000001
scalar w=0.0000001
matrix Cmet=Co+l*I(3)
```

(Υλοποίηση μεθόδου μεταφοράς)

Ο πίνακας συνδιακύμανσης S_{met} που προκύπτει από τη χρήση της μεθόδου της μεταφοράς είναι:

$$S_{met} = C_{met} = \begin{bmatrix} 475.37 & 231 & 475.86 \\ 231 & 117.24 & 161.06 \\ 475.86 & 161.06 & 1464.7 \end{bmatrix}$$

```
matrix SV=(sqrt(`var11'),0,0\0,sqrt(`var22'), 0 \ 0, 0,
sqrt(`var33'))
matrix R= inv(SV)*Co*inv(SV)
matrix symeigen Y r = R
mat list R
```

```
matrix D1=(r[1,1],0,0\ 0,r[1,2],0\0,0, 0.0000001) (Υλοποίηση
matrix BB=Y*D1*Y'                                 μεθόδου μέσω
matrix T=(1/sqrt(BB[1,1]) \ 1/sqrt(BB[2,2]) \      πίνακα
1/sqrt(BB[3,3]))                                   συσχέτισης)
matrix TT=T*T'
matrix Rnew=hadamard(BB,TT)
matrix symeigen Yr rnew = Rnew
matrix CovRnew=SV*Rnew*SV
mat list CovRnew
```

Ο πίνακας συνδιακύμανσης S_R που προκύπτει από τη χρήση της μεθόδου της μεταφοράς είναι:

$$S_R = Cov_{Rnew} = \begin{bmatrix} 468.16 & 222.98 & 467.72 \\ 222.98 & 110.03 & 161.17 \\ 467.72 & 161.17 & 1457.5 \end{bmatrix}$$

◊

3.2. Σύγκριση μεθόδων διαταραχής

Έστω οι $p \times p$ θετικά ημιορισμένοι πινάκες $S_{eig}, S_{met}, S_R = (\hat{s}_{ij})$, $i, j = 1, 2, \ldots, p$, όπως αυτοί υπολογίστηκαν στην Ενότητα 3.1, και $S = (s_{ij})$ ο πίνακας συνδιακύμανσης του συνόλου δεδομένων με πλήρεις παρατηρήσεις των $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ μεταβλητών.

Για τη σύγκριση των μεθόδων διαταραχής των ιδιοτιμών των παραπάνω συμμετρικών πινάκων θα χρησιμοποιηθούν:

1. νόρμα Frobenius,
2. κριτήριο διαφοράς με χρήση απόλυτων τιμών, και
3. γραμμική παλινδρόμηση.

3.2.1 Σύγκριση βάσει νόρμας Frobenius

Θεωρώντας S_{fr} τον πίνακα που ισούται με τη διαφορά του πίνακα S από τους S_{eig}, S_{met}, S_R, σύμφωνα με τον Ορισμό της νόρμας Frobenius στην (1.10) έχουμε:

$$|||\, S_{fr} \,|||_F = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{p} \left| \hat{s}_{ij} - s_{ij} \right|^2} \tag{3.6}$$

Ορίζουμε τον πραγματικό θετικό αριθμό:

$$fr = \frac{|||\, S_{fr} \,|||_F}{p^2} \tag{3.7}$$

3.2.3 Κριτήριο διαφοράς με χρήση απόλυτων τιμών

Ένας επιπρόσθετος τρόπος σύγκρισης των μεθόδων διαταραχής ιδιοτιμών είναι η χρήση ενός κριτηρίου διαφοράς το οποίο βασίζεται στις απόλυτες τιμές. Η ιδέα αυτή προέκυψε από [24], όπου χρησιμοποιείται ένα κριτήριο διαφοράς τετραγώνων προκειμένου να λάβει χώρα παρόμοια σύγκριση.

Ο τύπος που βασίζεται το κριτήριο της διαφοράς τετραγώνων είναι ο:

$$(SSD) = \sum_{i \le j} (\hat{s}_{ij} - s_{ij})^2$$

Ουσιαστικά, πρόκειται για ένα άθροισμα των τετραγωνικών διαφορών μεταξύ των αντίστοιχων $p(p+1)/2$ στοιχείων των συγκρινόμενων πινάκων.

Από την άλλη πλευρά , ο τύπος στον οποίο βασίζεται το κριτήριο διαφοράς με χρήση απόλυτων τιμών είναι:

$$(MSAD) = \frac{2}{p^2 + p} \sum_{i \le j} \left| \hat{s}_{ij} - s_{ij} \right| \tag{3.8}$$

Παράδειγμα 3.6

Έστω S ο πίνακας συνδιακύμανσης του συνόλου δεδομένων του Παραδείγματος 1.3 και οι πίνακες S_{eig}, S_{met}, S_R των Παραδειγμάτων 3.1, 3.2 και 3.4, αντίστοιχα. Να συγκριθούν βάσει νόρμας Frobenius και κριτηρίου διαφοράς με χρήση απόλυτων τιμών οι πινάκες.

Οι πίνακες συνδιακύμανσης που θα χρειαστούμε είναι οι ακόλουθοι:

$$S = \begin{bmatrix} 6 & 0.666 & -1.444 \\ 0.666 & 7.555 & -0.222 \\ -1.444 & -0.222 & 16.444 \end{bmatrix}, \quad S_{eig} = \begin{bmatrix} 6.6738 & 1.6425 & 2.4836 \\ 1.6425 & 9.7757 & 8.3794 \\ 2.4836 & 8.3794 & 7.3635 \end{bmatrix},$$

$$S_{met} = \begin{bmatrix} 10.4199 & 1.333 & 2.857 \\ 1.333 & 12.0589 & 10.22 \\ 2.857 & 10.22 & 8.9519 \end{bmatrix}, \quad S_R = \begin{bmatrix} 6.611 & 1.5509 & 2.175 \\ 1.5509 & 8.2500 & 6.4161 \\ 2.175 & 6.4161 & 5.143 \end{bmatrix}$$

i. Αρχικά, υπολογίζουμε τη νόρμα Frobenius του πίνακα $S_{fr} = S_{eig} - S$ με τη χρήση των σχέσεων (3.6) και (3.7).

$$|||S_{fr}|||_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\hat{s}_{ij} - s_{ij}\right|^2} = 16.3881$$

$$fr = \frac{|||S_{fr}|||_F}{p^2} = \frac{16.3881}{9} = 1.8209$$

Στη συνέχεια, εφαρμόζουμε το κριτήριο της διαφοράς με τη χρήση απόλυτων τιμών από τον (3.8).

$$(MSAD) = \frac{2}{p^2 + p}\sum_{i \le j}\left|\hat{s}_{ij} - s_{ij}\right| = \frac{7.3195}{6} = 1.219$$

ii. Επαναλαμβάνουμε την ίδια διαδικασία και για τον πίνακα S_{met}.

$$|||S_{fr}|||_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\hat{s}_{ij} - s_{ij}\right|^2} = 18.7689 \qquad fr = \frac{|||S_{fr}|||_F}{p^2} = \frac{17.06}{9} = 2.0854$$

$$(MSAD) = \frac{2}{p^2 + p}\sum_{i \le j}\left|\hat{s}_{ij} - s_{ij}\right| = \frac{31.8659}{6} = 5.31$$

iii. Επαναλαμβάνω την ίδια διαδικασία για τον πίνακα S_R.

$$|||S_{fr}|||_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\hat{s}_{ij} - s_{ij}\right|^2} = 15.6351 \qquad fr = \frac{|||S_{fr}|||_F}{p^2} = \frac{15.6351}{9} = 1.7372$$

$$(MSAD) = \frac{2}{p^2 + p}\sum_{i \le j}\left|\hat{s}_{ij} - s_{ij}\right| = \frac{23.749}{6} = 3.9581$$

Μέθοδοι	**Νόρμα Frobenius**	***fr***	**Κριτήριο διαφοράς με χρήση απόλυτων τιμών**
Μερική φασματική ανάλυση	16.3881	1.8209	1.219
Μεταφορά	18.7689	2.0854	5.31
Μέσω πινάκα συσχέτισης	15.6351	1.7372	3.9581

Πίνακας 3.1: Αποτελέσματα νόρμας Frobenius και κριτηρίου διαφοράς με τη χρήση απολύτων τιμών

Παρατηρώντας τον Πίνακα 3.1 αντιλαμβανόμαστε ότι τα αποτελέσματα από την εφαρμογή της νόρμας Frobenius και του κριτηρίου διαφοράς με χρήση απόλυτων τιμών, για το σύνολο δεδομένων που αναφέρθηκε στο Παράδειγμα 1.3, δεν έχουν συνάφεια. Άρα, δεν μπορεί να εξαχθεί ένα ασφαλές συμπέρασμα για το ποια μέθοδος θεωρείται βέλτιστη.

Ένας επιπρόσθετος λόγος, ο οποίος συνιστά ανασταλτικό παράγοντα εξαγωγής ασφαλούς συμπεράσματος είναι το γεγονός ότι το σύνολο δεδομένων που επεξεργαζόμαστε είναι σχετικά μικρό καθώς αποτελείται μόνο από 10 παρατηρήσεις.

◊

Παράδειγμα 3.7

Έστω S ο πίνακας συνδιακύμανσης που υπολογίστηκε στο Παράδειγμα 2.7 από το σύνολο δεδομένων του Παρατήματος Γ.1 και οι πίνακες S_{eig}, S_{met}, S_R του Παραδείγματος 3.5. Να συγκριθούν βάσει νόρμας Frobenius και κριτηρίου διαφοράς με χρήση απόλυτων τιμών οι πινάκες.

Οι πίνακες συνδιακύμανσης που θα χρειαστούμε είναι οι ακόλουθοι:

$$S = \begin{bmatrix} 380.73 & 145.26 & 385.61 \\ 145.26 & 101.29 & 208.45 \\ 385.61 & 208.45 & 1540.9 \end{bmatrix}, \quad S_{eig} = \begin{bmatrix} 469.86 & 227.95 & 475.64 \\ 227.95 & 115.51 & 161.45 \\ 475.64 & 161.45 & 1457.6 \end{bmatrix},$$

$$S_{met} = \begin{bmatrix} 475.37 & 231 & 475.86 \\ 231 & 117.24 & 161.06 \\ 475.86 & 161.06 & 1464.7 \end{bmatrix}, \quad S_R = \begin{bmatrix} 468.16 & 222.98 & 467.72 \\ 222.98 & 110.03 & 161.17 \\ 467.72 & 161.17 & 1457.5 \end{bmatrix}$$

i. Αρχικά, υπολογίζουμε τη νόρμα Frobenius του πίνακα $S_{fr} = S_{eig} - S$ με τη χρήση των σχέσεων (3.6) και (3.7).

$$||| S_{fr} |||_F = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{p} \left| \hat{s}_{ij} - s_{ij} \right|^2} = 222.2371$$

$$fr = \frac{||| S_{fr} |||_F}{p^2} = \frac{222.27}{9} = 24.693$$

Στη συνέχεια, εφαρμόζουμε το κριτήριο διαφοράς με τη χρήση απόλυτων τιμών από τον (3.8).

$$(MSAD) = \frac{2}{p^2 + p}\sum_{i \le j}\left|\hat{s}_{ij} - s_{ij}\right| = \frac{406.37}{6} = 67.678$$

ii. Επαναλαμβάνουμε την ίδια διαδικασία και για τον πίνακα S_{met}.

$$|||\, S_{fr}\, |||_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\hat{s}_{ij} - s_{ij}\right|^2} = 224.7265 \qquad fr = \frac{|||\, S_{fr}\, |||_F}{p^2} = \frac{1.87}{9} = 24.9696$$

$$(MSAD) = \frac{2}{p^2 + p}\sum_{i \le j}\left|\hat{s}_{ij} - s_{ij}\right| = \frac{410.17}{6} = 68.361$$

iii. Επαναλαμβάνω την ίδια διαδικασία για τον πίνακα S_R.

$$|||\, S_{fr}\, |||_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\hat{s}_{ij} - s_{ij}\right|^2} = 211.4513 \qquad fr = \frac{|||\, S_{fr}\, |||_F}{p^2} = \frac{211.45}{9} = 23.4946$$

$$(MSAD) = \frac{2}{p^2 + p}\sum_{i \le j}\left|\hat{s}_{ij} - s_{ij}\right| = \frac{286.68}{6} = 64.446$$

Μέθοδοι	**Νόρμα Frobenius**	***fr***	**Κριτήριο διαφοράς με χρήση απόλυτων τιμών**
Μερική φασματική ανάλυση	222.2371	24.693	67.678
Μεταφορά	224.7265	24.9696	68.361
Μέσω πινάκα συσχέτισης	211.4513	23.4946	64.446

Πίνακας 3.2: Αποτελέσματα νόρμας Frobenius και κριτηρίου διαφοράς με τη χρήση απόλυτων τιμών

Ο Πίνακας 3.2 μας οδηγεί στο συμπέρασμα ότι τα αποτελέσματα της σύγκρισης βάσει νόρμας Frobenius και βάσει κριτήριου διαφοράς με χρήση απόλυτων τιμών ταυτίζονται.

Ειδικότερα, το συμπέρασμα που εξάγεται από τις συγκεκριμένες σύγκρισης είναι ότι ο πίνακας συνδιακύμανσης που προκύπτει από τη μέθοδο μέσω πίνακα συσχέτισης είναι προσεγγιστικά πιο κοντά στον πίνακα συνδιακύμανσης που προέρχεται από το σύνολο δεδομένων χωρίς ελλιπείς τιμές. Ακολουθεί ο πίνακας συνδιακύμανσης που προκύπτει από τη μέθοδο της μερικής φασματικής ανάλυσης και τέλος αυτός που προκύπτει από τη μέθοδο της μεταφοράς. ◊

Παράδειγμα 3.8
Έστω

$$S = \begin{bmatrix} 0.000198 & 0.000489 & 0.000585 \\ 0.000489 & 0.002789 & 0.003718 \\ 0.000585 & 0.003718 & 0.006355 \end{bmatrix}$$

ο πίνακας συνδιακύμανσης του συνόλου δεδομένων του Παρατήματος Γ.2 (var1 – var3) και

$$S_{PW} = \begin{bmatrix} 0.000187 & 0.000492 & 0.000574 \\ 0.000492 & 0.002949 & 0.004316 \\ 0.000574 & 0.004316 & 0.006355 \end{bmatrix}$$

ένας πίνακας που προέκυψε με τη μέθοδο διαγραφής ανά ζεύγη. Υλοποιώντας τμήμα του κώδικα του Παραρτήματος Β με $\varepsilon = 10^{-7}$ υπολογίζονται οι πίνακες:

$$S_{eig} = \begin{bmatrix} 0.000194 & 0.000481 & 0.000581 \\ 0.000481 & 0.002967 & 0.004305 \\ 0.000581 & 0.004305 & 0.006362 \end{bmatrix}, \quad S_{met} = \begin{bmatrix} 0.000219 & 0.000492 & 0.000574 \\ 0.000492 & 0.002981 & 0.004316 \\ 0.000574 & 0.004316 & 0.006387 \end{bmatrix},$$

$$S_R = \begin{bmatrix} 0.000187 & 0.000489 & 0.000574 \\ 0.000489 & 0.002949 & 0.00427 \\ 0.000574 & 0.00427 & 0.006355 \end{bmatrix}$$

Να συγκριθούν βάσει νόρμας Frobenius και κριτηρίου διαφοράς με χρήση απόλυτων τιμών οι πινάκες.

i. Αρχικά, υπολογίζουμε τη νόρμα Frobenius του πίνακα $S_{fr} = S_{eig} - S$ με τη χρήση των σχέσεων (3.6) και (3.7).

$$|||\, S_{fr} \,|||_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\hat{s}_{ij} - s_{ij}\right|^2} = 8.4914 \cdot 10^{-4}$$

$$fr = \frac{|||\, S_{fr} \,|||_F}{p^2} = \frac{8.4914 \cdot 10^{-4}}{9} = 9.4349 \cdot 10^{-5}$$

Στη συνέχεια, εφαρμόζουμε το κριτήριο διαφοράς με τη χρήση απόλυτων τιμών από τον (3.8).

$$(MSAD) = \frac{2}{p^2 + p}\sum_{i \le j}\left|\hat{s}_{ij} - s_{ij}\right| = \frac{0.788 \cdot 10^{-3}}{6} = 0.1313 \cdot 10^{-3}$$

ii. Επαναλαμβάνουμε την ίδια διαδικασία και για τον πίνακα S_{met}.

$$|||\,S_{fr}\,|||_F=\sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\hat{s}_{ij}-s_{ij}\right|^2}=8.6821\cdot 10^{-4}$$

$$fr=\frac{|||\,S_{fr}\,|||_F}{p^2}=\frac{8.6821\cdot 10^{-4}}{9}=9.6468\cdot 10^{-5}$$

$$(MSAD)=\frac{2}{p^2+p}\sum_{i\leq j}\left|\hat{s}_{ij}-s_{ij}\right|=\frac{0.857\cdot 10^{-3}}{6}=0.142\cdot 10^{-3}$$

iii. Επαναλαμβάνω την ίδια διαδικασία για τον πίνακα S_R.

$$|||\,S_{fr}\,|||_F=\sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\hat{s}_{ij}-s_{ij}\right|^2}=7.971\cdot 10^{-4}$$

$$fr=\frac{|||\,S_{fr}\,|||_F}{p^2}=\frac{7.971\cdot 10^{-4}}{9}=8.8567\cdot 10^{-5}$$

$$(MSAD)=\frac{2}{p^2+p}\sum_{i\leq j}\left|\hat{s}_{ij}-s_{ij}\right|=\frac{0.734\cdot 10^{-3}}{6}=0.122\cdot 10^{-3}$$

Μέθοδοι	**Νόρμα Frobenius**	***fr***	**Κριτήριο διαφοράς με χρήση απόλυτων τιμών**
Μερική φασματική ανάλυση	$8.4914\cdot 10^{-4}$	$9.4349\cdot 10^{-5}$	$0.1313\cdot 10^{-3}$
Μεταφορά	$8.6821\cdot 10^{-4}$	$9.6468\cdot 10^{-5}$	$0.142\cdot 10^{-3}$
Μέσω πινάκα συσχέτισης	$7.971\cdot 10^{-4}$	$8.8567\cdot 10^{-5}$	$0.122\cdot 10^{-3}$

Πίνακας 3.3: Αποτελέσματα νόρμας Frobenius και κριτηρίου διαφοράς με τη χρήση απόλυτων τιμών

Παρατηρώντας τον Πίνακα 3.3 καταλήγουμε στο συμπέρασμα ότι ο πίνακας συνδιακύμανσης που προκύπτει από τη μέθοδο μέσω πίνακα συσχέτισης είναι προσεγγιστικά πιο κοντά στον πίνακα συνδιακύμανσης που προέρχεται από το σύνολο δεδομένων χωρίς ελλιπείς τιμές. Ακολουθεί ο πίνακας συνδιακύμανσης που προκύπτει από τη μέθοδο της μερικής φασματικής ανάλυσης και τέλος αυτός που προκύπτει από τη μέθοδο της μεταφοράς.

◊

Παράδειγμα 3.9

Έστω

$$S = \begin{bmatrix} 12.4189 & 4.80355 & 16.5137 \\ 4.80355 & 4.08789 & 6.81312 \\ 16.5137 & 6.81312 & 23.3602 \end{bmatrix}$$

ο πίνακας συνδιακύμανσης του συνόλου δεδομένων του Παρατήματος Γ.2 (var3 – var5) και

$$S_{PW} = \begin{bmatrix} 12.957 & 3.9103 & 17.026 \\ 3.9103 & 2.8846 & 5.6253 \\ 17.026 & 5.6253 & 22.195 \end{bmatrix}$$

ένας πίνακας που προέκυψε με τη μέθοδο διαγραφής ανά ζεύγη. Υλοποιώντας τμήμα του κώδικα του Παραρτήματος Β με $\varepsilon = 10^{-7}$ υπολογίζονται οι πίνακες:

$$S_{eig} = \begin{bmatrix} 13.028 & 3.9272 & 16.967 \\ 3.9272 & 2.8887 & 5.6113 \\ 16.967 & 5.6113 & 22.243 \end{bmatrix}, \quad S_{met} = \begin{bmatrix} 13.079 & 3.9103 & 17.025 \\ 3.9103 & 3.0071 & 5.6253 \\ 17.025 & 5.6253 & 22.318 \end{bmatrix},$$

$$S_R = \begin{bmatrix} 12.957 & 3.9059 & 16.899 \\ 3.9059 & 2.8846 & 5.6103 \\ 16.899 & 5.6103 & 22.195 \end{bmatrix}$$

Να συγκριθούν βάσει νόρμας Frobenius και κριτηρίου διαφοράς με χρήση απόλυτων τιμών οι πινάκες.

i. Αρχικά, υπολογίζουμε τη νόρμα Frobenius του πίνακα $S_{fr} = S_{eig} - S$ με τη χρήση των σχέσεων (3.6) και (3.7).

$$|||\, S_{fr}\, |||_F = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{p} \left| \hat{s}_{ij} - s_{ij} \right|^2} = 2.8094$$

$$fr = \frac{|||\, S_{fr}\, |||_F}{p^2} = \frac{2.8094}{9} = 0.3122$$

Στη συνέχεια, εφαρμόζουμε το κριτήριο διαφοράς με τη χρήση απόλυτων τιμών από τον (3.8). .

$$(MSAD) = \frac{2}{p^2 + p}\sum_{i \le j}\left|\hat{s}_{ij} - s_{ij}\right| = \frac{5.4573}{6} = 0.9095$$

ii. Επαναλαμβάνουμε την ίδια διαδικασία και για τον πίνακα S_{met}.

$$|||\, S_{fr}\, |||_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\hat{s}_{ij} - s_{ij}\right|^2} = 2.7623 \qquad fr = \frac{|||\, S_{fr}\, |||_F}{p^2} = \frac{2.7623}{9} = 0.3069$$

$$(MSAD) = \frac{2}{p^2 + p}\sum_{i \le j}\left|\hat{s}_{ij} - s_{ij}\right| = \frac{5.3755}{6} = 0.8959$$

iii. Επαναλαμβάνω την ίδια διαδικασία για τον πίνακα S_R.

$$|||\, S_{fr}\, |||_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\hat{s}_{ij} - s_{ij}\right|^2} = 2.8092 \qquad fr = \frac{|||\, S_{fr}\, |||_F}{p^2} = \frac{2.8092}{9} = 0.3121$$

$$(MSAD) = \frac{2}{p^2 + p}\sum_{i \le j}\left|\hat{s}_{ij} - s_{ij}\right| = \frac{5.3874}{6} = 0.8979$$

Μέθοδοι	**Νόρμα Frobenius**	***fr***	**Κριτήριο διαφοράς με χρήση απόλυτων τιμών**
Μερική φασματική ανάλυση	2.8094	0.3122	0.9095
Μεταφορά	2.7623	0.3069	0.8959
Μέσω πινάκα συσχέτισης	2.8092	0.3121	0.8979

Πίνακας 3.4: Αποτελέσματα νόρμας Frobenius και κριτηρίου διαφοράς με τη χρήση απόλυτων τιμών

Το συμπέρασμα που εξάγεται από τον Πίνακα 3.4 είναι πως ο πίνακας συνδιακύμανσης που προκύπτει από τη μέθοδο της μεταφοράς είναι προσεγγιστικά πιο κοντά στον πίνακα συνδιακύμανσης που προέρχεται από το σύνολο δεδομένων χωρίς ελλιπή στοιχεία. Ακολουθεί ο πίνακας συνδιακύμανσης που προκύπτει από τη μέθοδο μέσω πίνακα συσχέτισης και τέλος αυτός που προκύπτει από τη μέθοδο της μερικής φασματικής ανάλυσης. ◊

Σχόλιο 3.2

Τα αποτελέσματα από την εφαρμογή της νόρμας Frobenius και του κριτηρίου διαφοράς με τη χρήση απόλυτων τιμών στα Παραδείγματα 3.7 - 3.9 σηματοδοτούν την ενδεχόμενη εξάρτηση της κατάταξης των διαταραγμένων μεθόδων από το pvalue των μεταβλητών. Αξίζει να σημειωθεί ότι η παρατήρηση αυτή χρήζει περαιτέρω μελέτης και εμβάθυνσης.

3.2.3 Σύγκριση βάσει γραμμικής παλινδρόμησης

Στην Ενότητα 1.5 κατέστει σαφής η έννοια της γραμμικής παλινδρόμησης καθώς και ο τρόπος υπολογισμού των αγνώστων παραμέτρων b της σχέσης (1.26). Στο κεφάλαιο αυτό θα χρησιμοποιηθεί η έννοια της γραμμικής παλινδρόμησης προκειμένου να λάβει χώρα η σύγκριση των μεθόδων των παρατηρήσεων με ελλιπή δεδομένα, της μερικής φασματικής ανάλυσης, της μεταφοράς και της μεθόδου μέσω πίνακα συσχέτισης. Η μεθοδολογία που θα ακολουθηθεί προκειμένου να τελεσφορήσει ο υπολογισμός των αγνώστων παραμέτρων b είναι ίδια με αυτή που παρουσιάζεται στην Πρόταση 1.14. Η μόνη διαφορά έγκειται στο γεγονός ότι η άγνωστη παράμετρος b_0 δεν είναι δυνατόν να υπολογιστεί για τις μεθόδους της μερικής φασματικής ανάλυσης, της μεταφοράς και της μεθόδου μέσω πίνακα συσχέτισης καθώς το σύνολο δεδομένων από το οποίο προέρχονται αυτοί οι πίνακες, έχει ελλιπή δεδομένα.

Παράδειγμα 3.10

Υπολογισμός των αγνώστων παραμέτρων $\boldsymbol{b}$ της σχέσης (1.26) για το σύνολο δεδομένων του Παραδείγματος 2.6 για τις περιπτώσεις:

i. της μερικής φασματικής ανάλυσης
ii. της μεταφοράς και
iii. της μεθόδου μέσω πίνακα συσχέτισης.

Το διάνυσμα μέσης τιμής ισούται με: $\boldsymbol{\mu} = \begin{bmatrix} 2.8 \\ 2.75 \\ 3 \end{bmatrix}$

i. Ο πίνακας συνδιακύμανσης που προκύπτει από τη μέθοδο της μερικής φασματικής ανάλυσης είναι ο: $S_{eig} = \begin{bmatrix} 6.01 & 0.02 & -2.23 \\ 0.02 & 7.34 & -8.27 \\ -2.23 & -8.27 & 10.14 \end{bmatrix}$

Με τη χρήση των σχέσεων (1.22) υπολογίζουμε τους συντελεστές b.

$$\boldsymbol{b}_{eig} = S_{22}^{-1}\boldsymbol{S}_{12}^{T} = \begin{bmatrix} 7.34 & -8.27 \\ -8.27 & 10.14 \end{bmatrix}^{-1} \begin{bmatrix} 0.02 \\ -2.23 \end{bmatrix} = \begin{bmatrix} -3.02 \\ -2.68 \end{bmatrix}$$

Το $\boldsymbol{b}_0$ όπως προαναφέρθηκε παραπάνω δεν είναι εφικτό να υπολογιστεί. Άρα η συνάρτηση εκτιμώμενης παλινδρόμησης είναι η:

$$\boldsymbol{b}_0 + \boldsymbol{b}_{eig} = \boldsymbol{b}_0 + 3.02\boldsymbol{X}_1 + 2.68\boldsymbol{X}_2$$

ii. Ο πίνακας συνδιακύμανσης που προκύπτει από τη μέθοδο της μεταφοράς είναι ο: $S_{met} = \begin{bmatrix} 10.3 & 1 & 2.85 \\ 1 & 11.95 & 10 \\ 2.85 & 10 & 8.84 \end{bmatrix}$

Με τη χρήση των σχέσεων (1.22) υπολογίζουμε τους συντελεστές $\boldsymbol{b}$.

$$\boldsymbol{b}_{met} = S_{22}^{-1}\boldsymbol{S}_{12}^{T} = \begin{bmatrix} 11.95 & 10 \\ 10 & 8.84 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2.85 \end{bmatrix} = \begin{bmatrix} -3.49 \\ 4.27 \end{bmatrix}$$

Το $\boldsymbol{b}_0$ όπως προαναφέρθηκε παραπάνω δεν είναι εφικτό να υπολογιστεί. Άρα η συνάρτηση εκτιμώμενης παλινδρόμησης είναι η:

$$\boldsymbol{b}_0 + \boldsymbol{b}_{met} = \boldsymbol{b}_0 + 3.49\boldsymbol{X}_1 - 4.27\boldsymbol{X}_2$$

iii. Ο πίνακας συνδιακύμανσης που προκύπτει από τη μέθοδο μέσω πίνακα συνδιακύμανσης είναι ο: $S_R = \begin{bmatrix} 6.6 & 0 & 0 \\ 0 & 8.24 & 0 \\ 0 & 0 & 5.15 \end{bmatrix}$

Με τη χρήση της σχέσης (1.22) υπολογίζουμε τους συντελεστές $\boldsymbol{b}$.

$$\boldsymbol{b}_R = S_{22}^{-1}\boldsymbol{S}_{12}^{T} = \begin{bmatrix} 8.24 & 0 \\ 0 & 5.15 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Το $\boldsymbol{b}_0$ όπως προαναφέρθηκε παραπάνω δεν είναι εφικτό να υπολογιστεί. Άρα η συνάρτηση εκτιμώμενης παλινδρόμησης είναι η:

$$\boldsymbol{b}_0 + \boldsymbol{b}_R = \boldsymbol{b}_0 - 0X_1 - 0X_2$$

Μέθοδοι	b_1	b_2
Μερική φασματική ανάλυση	–3.02	–2.68
Μεταφορά	–3.49	4.27
Μέσω πινάκα συσχέτισης	0	0

Πίνακας 3.5:Οι συντελεστές γραμμικής παλινδρόμησης για τις μεθόδους διαταραχής

Τα αποτελέσματα από την εφαρμογή της γραμμικής παλινδρόμησης στο Παράδειγμα 3.1 δεν μπορούν να αξιοποιηθούν προκειμένου να εξαχθεί ένα ασφαλές συμπέρασμα για το ποια μέθοδος θεωρείται βέλτιστη.

Ο κύριος λόγος, ο οποίος συνιστά ανασταλτικό παράγοντα εξαγωγής ασφαλούς συμπεράσματος είναι το γεγονός ότι το σύνολο δεδομένων που επεξεργαζόμαστε είναι σχετικά μικρό καθώς αποτελείται μόνο από 10 παρατηρήσεις. ◊

Παράδειγμα 3.8

Αξιοποιώντας το σύνολο δεδομένων του Παραρτήματος Γ.1 και τμήματα του κώδικα του Παραρτήματος Β που παρατίθενται παρακάτω υπολογίζουμε τις άγνωστες παραμέτρους $\boldsymbol{b}$ της σχέσης (1.26) για τις περιπτώσεις:

1. της μερικής φασματικής ανάλυσης
2. της μεταφοράς και
3. της μεθόδου μέσω πίνακα συσχέτισης.

```
matrix Bn1 = Cnew1[1,2..3]
matrix Bn2 = Cnew1[2..3,2..3]
matrix bnega2 = Bn1*invsym(Bn2)
```

Οι άγνωστοι παράμετροι $\boldsymbol{b}$ της σχέσης (1.26) που προκύπτουν από τη χρήση της μεθόδου της μερικής φασματικής ανάλυσης είναι οι:

$$\boldsymbol{b}_{nega2} = \boldsymbol{b}_{eig} = \begin{bmatrix} 2.0369282 \\ 0.00170633 \end{bmatrix}$$

```
matrix L12=L[1,2..3]
matrix L22=L[2..3,2..3]
matrix bm=L12*invsym(L22)
mat list L
```

Οι άγνωστοι παράμετροι $\boldsymbol{b}$ της σχέσης (1.26) που προκύπτουν από τη χρήση της μεθόδου της μεταφοράς είναι οι:

$$\boldsymbol{b}_{m} = \boldsymbol{b}_{met} = \begin{bmatrix} 2.0369281 \\ 0.00170633 \end{bmatrix}$$

```
matrix CR12=CovRnew[1,2..3]
matrix CR22=CovRnew[2..3,2..3]
matrix bcorrelm=CR12*invsym(CR22)
mat list CovRnew
```

Οι άγνωστοι παράμετροι $\boldsymbol{b}$ της σχέσης (1.21) που προκύπτουν από τη χρήση της μεθόδου μέσω πίνακα συσχέτισης είναι οι:

$$\boldsymbol{b}_{correlm} = \boldsymbol{b}_R = \begin{bmatrix} 2.0715518 \\ -0.00343424 \end{bmatrix}$$

Μέθοδοι	b_1	b_2
Μερική φασματική ανάλυση	2.0369282	0.00170633
Μεταφορά	2.0369281	0.00170633
Μέσω πινάκα συσχέτισης	2.0715518	-0.00343424

Πίνακας 3.6: Οι συντελεστές γραμμικής παλινδρόμησης για τις μεθόδους διαταραχής

Εύστοχα παρατηρώντας τον Πίνακα 3.6 αντιλαμβανόμαστε ότι τα $\boldsymbol{b}$ που προέκυψαν με τη χρήση του πίνακα συνδιακύμανσης από τη μέθοδο της ειδικής φασματικής ανάλυσης και αυτά από τη μέθοδο της μεταφοράς είναι σχεδόν ίσα, δηλαδή $\boldsymbol{b}_{eig} \approx \boldsymbol{b}_{met}$. Τα $\boldsymbol{b}_R$ ωστόσο είναι λίγο μεγαλύτερα.

Συνεπώς, το συμπέρασμα που εξάγεται από τη συγκεκριμένη σύγκριση είναι πως οι πίνακες συνδιακύμανσης που προκύπτουν από τις μεθόδους της μερικής φασματικής ανάλυσης και της μεταφοράς είναι προσεγγιστικά πιο κοντά στο πίνακα συνδιακύμανσης που προέρχεται από το σύνολο δεδομένων του Παραρτήματος Γ.1 σε σχέση με τον πίνακα συνδιακύμανσης που προκύπτει από τη μέθοδο μέσω πίνακα συσχέτισης. ◊

Παράδειγμα 3.9

Αξιοποιώντας το σύνολο δεδομένων του Παραρτήματος Γ.2 (var1 – var3) και τμήματα του κώδικα του Παραρτήματος Β υπολογίζουμε τις άγνωστες παραμέτρους $\boldsymbol{b}$ της σχέσης (1.21) για τις περιπτώσεις:

i. της μερικής φασματικής ανάλυσης
ii. της μεταφοράς και
iii. της μεθόδου μέσω πίνακα συσχέτισης.

Οι άγνωστοι παράμετροι $\boldsymbol{b}$ της σχέσης (1.21) που προκύπτουν από τη χρήση της μεθόδου της μερικής φασματικής ανάλυσης είναι οι:

$$\boldsymbol{b}_{nega2} = \boldsymbol{b}_{eig} = \begin{bmatrix} 1.6088654 \\ -0.97546895 \end{bmatrix}$$

Οι άγνωστοι παράμετροι $\boldsymbol{b}$ της σχέσης (1.21) που προκύπτουν από τη χρήση της μεθόδου της μεταφοράς είναι οι:

$$\boldsymbol{b}_{m} = \boldsymbol{b}_{met} = \begin{bmatrix} 1.5874827 \\ -0.96161396 \end{bmatrix}$$

Οι άγνωστοι παράμετροι $\boldsymbol{b}$ της σχέσης (1.21) που προκύπτουν από τη χρήση της μεθόδου μέσω πίνακα συσχέτισης είναι οι:

$$\boldsymbol{b}_{correlm} = \boldsymbol{b}_{R} = \begin{bmatrix} 1.2741958 \\ -0.74851765 \end{bmatrix}$$

Μέθοδοι	b_1	b_2
Ειδική φασματική ανάλυση	1.6088654	-0.97546895
Μεταφορά	1.5874827	-0.96161396
Μέσω πινάκα συσχέτισης	1.2741958	-0.74851765

Πίνακας 3.7: Οι συντελεστές γραμμικής παλινδρόμησης για τις μεθόδους διαταραχής

Παρατηρώντας τον Πίνακα 3.7 καθίσταται σαφές ότι η μέθοδος μέσω πίνακα συσχέτισης είναι προσεγγιστικά πιο κοντά στον πίνακα συνδιακύμανσης που προέρχεται από το σύνολο δεδομένων του Παραρτήματος Γ.2 (var1 – var3) σε σχέση με τους πίνακες συνδιακύμανσης που προκύπτουν από τις μεθόδους της μερικής φασματικής ανάλυσης και της μεταφοράς. ◊

Παράδειγμα 3.10

Αξιοποιώντας το σύνολο δεδομένων του Παραρτήματος Γ.2 (var4 – var6) και τμήματα του κώδικα του Παραρτήματος Β υπολογίζουμε τις άγνωστες παραμέτρους $\boldsymbol{b}$ της σχέσης (1.21) για τις περιπτώσεις:

i. της μερικής φασματικής ανάλυσης

ii. της μεταφοράς και

iii. της μεθόδου μέσω πίνακα συσχέτισης.

Οι άγνωστοι παράμετροι $\boldsymbol{b}$ της σχέσης (1.21) που προκύπτουν από τη χρήση της μεθόδου της μερικής φασματικής ανάλυσης είναι οι:

$$\boldsymbol{b}_{nega2} = \boldsymbol{b}_{eig} = \begin{bmatrix} \text{-}0.11102335 \\ 0.76522978 \end{bmatrix}$$

Οι άγνωστοι παράμετροι $\boldsymbol{b}$ της σχέσης (1.21) που προκύπτουν από τη χρήση της μεθόδου της μεταφοράς είναι οι:

$$\boldsymbol{b}_{m} = \boldsymbol{b}_{met} = \begin{bmatrix} \text{-}0.11102333 \\ 0.76522977 \end{bmatrix}$$

Οι άγνωστοι παράμετροι $\boldsymbol{b}$ της σχέσης (1.21) που προκύπτουν από τη χρήση της μεθόδου μέσω πίνακα συσχέτισης είναι οι:

$$\boldsymbol{b}_{correlm} = \boldsymbol{b}_{R} = \begin{bmatrix} \text{-}0.10561589 \\ 0.75907745 \end{bmatrix}$$

Μέθοδοι	b_1	b_2
Ειδική φασματική ανάλυση	-0.11102335	0.76522978
Μεταφορά	-0.11102333	0.76522977
Μέσω πινάκα συσχέτισης	-0.10561589	0.75907745

Πίνακας 3.8: Οι συντελεστές γραμμικής παλινδρόμησης για τις μεθόδους διαταραχής

Παρατηρώντας τον Πίνακα 3.8 εξάγεται το συμπέρασμα ότι η μέθοδος μέσω πίνακα συσχέτισης είναι προσεγγιστικά πιο κοντά στον πίνακα συνδιακύμανσης που προέρχεται από το σύνολο δεδομένων του Παραρτήματος Γ.2 (var1 – var3) σε σχέση με τους πίνακες συνδιακύμανσης που προκύπτουν από τις μεθόδους της μερικής φασματικής ανάλυσης και της μεταφοράς. ◊

Σχόλια 3.2

Τα αποτελέσματα από την εφαρμογή της γραμμικής παλινδρόμησης στα Παραδείγματα 3.8 - 3.10 σηματοδοτούν την ενδεχόμενη εξάρτηση της κατάταξης των διαταραγμένων μεθόδων από το μέγεθος του εκάστοτε συνόλου δεδομένων. Αξίζει να σημειωθεί ότι η παρατήρηση αυτή χρήζει περαιτέρω μελέτης και εμβάθυνσης. □

Συμπεράσματα

Η μελέτη της διαταραχής ιδιοτιμών συμμετρικών πινάκων ήταν το θέμα που μας απασχόλησε στην παρούσα εργασία. Αρχικά, παρουσιάστηκαν τα σημαντικότερα θεωρήματα και πορίσματα της πολυμεταβλητής στατιστικής, τα οποία συμβάλλουν στην εκτίμηση των στατιστικών πινάκων.

Στη συνέχεια, παρατέθηκαν μέθοδοι υπολογισμού πινάκων συνδιακύμανσης όταν το σύνολο δεδομένων από το οποίο προέρχονται περιέχει και ελλιπή δεδομένα εκτός από πλήρεις εγγραφές. Ιδιαίτερο ενδιαφέρον είχαν τα αποτελέσματα της μεθόδου διαγραφής ανά ζεύγη, τα οποία κάποιες φορές δεν επαληθεύουν την ιδιότητα ενός πίνακα συνδιακύμανσης, γεγονός που δημιουργεί σοβαρό πρόβλημα σε διάφορες πολυμεταβλητές τεχνικές ανάλυσης δεδομένων, όπως για παράδειγμα, στην παλινδρόμηση (regression) και στην ανάλυση κατά παράγοντες (factor analysis).

Ο ένας από τους δυο κώδικες σε stata που υλοποιήθηκαν στο πλαίσιο της παρούσας πτυχιακής εργασίας υπολογίζει πίνακες συνδιακύμανσης με τη χρήση των μεθόδων των παρατηρήσεων με ελλιπή δεδομένα και της διαγραφής ανά ζεύγη.

Ο άλλος κώδικας σε stata υλοποιεί τις μεθόδους διαταραχής των ιδιοτιμών συμμετρικών πινάκων. Ουσιαστικά, οι μέθοδοι που υλοποιούνται είναι αυτές της μερικής φασματικής ανάλυσης, της μεταφοράς και μέσω πίνακα συσχέτισης.

Τέλος, συγκρίνοντας τις μεθόδους διαταραχής τόσο με τη νόρμα Frobenius και με το κριτήριο διαφοράς με τη χρήση απόλυτων τιμών καθώς και με τη μέθοδο της γραμμικής παλινδρόμησης, καταλήξαμε στο συμπέρασμα ότι η κατάταξη των μεθόδων διαταραχής ενδεχομένως να εξαρτάται από το μέγεθος του συνόλου δεδομένων που μελετάται και από το pvalue των μεταβλητών.

Τα σύνολα δεδομένων με ελλιπή δεδομένα παρουσιάζονται σε ποικίλες ερευνητικές περιοχές, γεγονός που δημιουργεί προβλήματα. Αυτά αποτελούν πεδίο έρευνας, το προϊόν της οποίας εμπλουτίζει καθημερινά τη διεθνή αρθρογραφία. Έτσι η προαναφερθείσα θεωρία η σχετική με τις μεθόδους διαταραχής καθώς και τα συμπεράσματα που εξήχθησαν, είναι αρκετά χρήσιμα και χρήζουν περαιτέρω μελέτης και εμβάθυνσης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Pantelis G. Bagos, Theodorore D. Liakopoulos, A Multipoint Method for Meta-Analysis of Genetic Association Studies, *Genetic Epidemiology*, **34**, (2010), 702-715.
2. Sean Borman, *The Expectation Maximization Algorithm A short tutorial*, (2004).
3. Rufus Lynn Carter, Solutions for Missing Data in Structural Equation Modeling, *Research & Practice in Assessment*, **1**(1), (2006).
4. T. Carter, R. Tapia and A. Papakonstantinou, *Linear Algebra-An Intoduction to Linear Algebra for Pre-Calculus Students*, Rice University, 1995.
5. Christophe Couvreur, The EM Algorithm: A Guided Tour, *Proceedings of the 2nd IEEE European Workshop on Computationaly Intensive Methods in Control and Signal Processing, Pragues, Czech Rep.*, (August, 1996), 115-120.
6. Χ. Δέδες, *Μη αρνητικοί πίνακες*, Πτυχιακή εργασία, Λαμία, 2011.
7. Γ. Δονάτος και Μ. Αδάμ, *Γραμμική Άλγεβρα. Θεωρία και εφαρμογές*, Εκδόσεις Guttenberg, Αθήνα, 2008.
8. Γ. Δονάτος και Β. Χομπάς, *Στατιστικές Μέθοδοι*, Εκδόσεις Αντ. Ν. Σακκούλα, Αθήνα-Κομοτηνή, 1988.
9. Frank Dellaert, The Expectation Maximization Algorithm, *GVU Center*, Technical Report number GIT-GVU-02-20, (2002).
10. Brian S. Everitt and Graham Dunn, *Applied Multivariate Data Analysis*, Hodder Arnold, Second Edition, London, 2001.
11. Francis Galton, Regression towards mediocrity in hereditary stature, *The Journal of the Anthropological Institute of Great Britain and Ireland*, **15**, (1886), 246–263.
12. G. Golub and C. V. Loan, *Matrix Computations*, The Johns Hopkins University Prress, 3rd edition, Baltimore, 1996.
13. John W. Graham, Missing Data Analysis: Making It Work in the Real World, *Annual Review of Phycology*, (2009), 549-576.
14. R. A. Horn and C. R. Johnson, Matrix analysis, *Cambridge University Press*, New York, 2005.

15. Richard A. Johnson and Dean W. Wichern, *Applied Multivariate Statistical Analysis*, Pearson International Edition, Sixth Edition, United States of America, 2007.
16. Ν. Καδιανάκης και Σ. Καρανάσιος, *Γραμμική Άλγεβρα και Στοιχεία Αναλυτικής Γεωμετρίας*, Αθήνα, 1989.
17. Dirk L. Knol and Jos M. F. Xen Berge, Least-Squares Approximation of an Improper Correlation Matrix by a Proper one, Psychometrica, **54**(1), (1989), 53-61.
18. Φ. Κολυβά - Μαχαίρα, Ε. Μπόρα - Σέντα, *Στατιστική, Θεωρία, Εφαρμογές*, Εκδόσεις Ζήτη, Θεσσαλονίκη, 1998.
19. Σ. Κουνιάς, Φ. Κολυβά - Μαχαίρα, Κ. Μπαγιάτης, Ε. Μπόρα - Σέντα, Εισαγωγή στη Στατιστική, Εκδόσεις Γιαχούδη, Θεσσαλονίκη, 1985.
20. R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Second Edition, John Wiley & Sons, New Jersey, 2002.
21. C. Meyer, *Matrix analysis and applied linear algebra*, SIAM Press, 2000.
22. David A. Penn, Financial Well-Being in an Urban Setting: An Application of Multiple Imputation, *Department of Economics and Finance Working Paper Series*, (2005).
23. Peter J. Rousseuw, Geert Molenberghs, Transformation of non positive semidefinite correlation matrices, *Communications in Statistics-Theory and Methonds*, **22**(4), (1993), 965-984.
24. N. C. Schwertman, D. M. Allen, Smoothing an indefinite variance-covariance matrix, *Journal of Statistical Computation and Simulation*, **9**(3), (1979), 183-194.
25. Murray R. Spiegel, Larry J. Stephens, *Στατιστική*, $3^{η}$ έκδοση, Εκδόσεις Τζιόλα, Θεσσαλονίκη, 2000.
26. Werner Wothke, SmallWaters Corp., Longitudinal and multi-group modeling with missing data, *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples*, Mahwah, NJ: Lawrence Erlbaum, (1998).

Χρήσιμα Links:

27. Enrico Schumann, *'Repairing' an Indefinite Correlation Matrix*, Comisef Wiki, http://comisef.wikidot.com/tutorial:repairingcorrelation.
28. http://archive.ics.uci.edu/ml/
29. http://archive.ics.uci.edu/ml/datasets/Liver+Disorders
30. http://archive.ics.uci.edu/ml/datasets /Breast+Cancer+Wisconsin+(Diagnostic)

ΠΑΡΑΡΤΗΜΑ Α

Στη συνέχεια παρατίθεται σε Stata ο κώδικας που αναπτύχθηκε στην παρούσα πτυχιακή εργασία για τον υπολογισμό πίνακα συνδιακύμανσης σύμφωνα με τη θεωρία που αναπτύχθηκε στην παράγραφο 1.4.3 (Ορισμός 1.18), στην παράγραφο 2.1 (μέθοδος παρατηρήσεων με ελλιπή δεδομένα), στην παράγραφο 2.3 (μέθοδος διαγραφής ανά ζεύγη). Επίσης, υπολογίζεται πίνακας συνδιακύμανσης με σκοπό να υπολογιστούν συντελεστές γραμμικής παλινδρόμησης ενός πολυμεταβλητού μοντέλου.

```
set more off
use "C:\data\stata\cancer_breast.dta", clear

correlate var1-var3,cov           (Υπολογισμός του πίνακα συνδιακύμανσης
matrix Cf = r(C)                  από σύνολο με πλήρη δεδομένα, Ορισμός
                                  1.18)

matrix W12 = Cf[1,2..3]           (Διαμέριση πίνακα Cf και υπολογισμός
matrix W22 = Cf[2..3,2..3]         συντελεστών γραμμικής παλινδρόμησης,
matrix bfull = W12*invsym(W22)     Ενότητα 1.5)


matrix Bhat=(0,0)                  (Αρχικοποίηση πινάκων)
matrix Bpos=(0,0)
matrix BhatEb2=(0,0)
matrix BposEb2=(0,0)

scalar pos=0                      (Αρχικοποίηση μεταβλητών)
scalar a=1
scalar b=3
scalar m=0
scalar n=15000                    (Μέγιστη τιμή επιθυμητών επαναλήψεων)
while(a==1){
use "C:\data\stata\cancer_breast.dta",clear
```

```
local nmiss=556

scalar m=m+1                        (Δήλωση μετρητή που μετρά το πλήθος
                                     των επαναλήψεων του προγράμματος)

if(m==n){
scalar a=2                          (Έξοδος από το while)
}

forvalues j=1(1)`nmiss' {            (Δημιουργία ελλιπών στοιχείων
         local z=uniform()            στο σύνολο δεδομένων)
         local y=`z' *_N
         local y=ceil(`y')
         local z=uniform()
         local x=`z' *3
         local x=ceil(`x')
         replace var`x'=. in `y'
 }

forvalues x=1/3 {                     (Δημιουργία πίνακα συνδιακύμανσης με
  forvalues y=1/3{                     μέθοδο διαγραφής ανά ζεύγη, Ενότητα
   correlate var`x' var`y',cov         2.3)
     local var`x'`y'= r(cov_12)
  }
}

 matrix Co =(`var11',`var12', `var13' \ `var21',`var22',
`var23' \ `var31',`var32' ,`var33' )

matrix symeigen X v = Co            (Υπολογισμός ιδιοδιανυσμάτων και
                                     ιδιοτιμών πίνακα συνδιακύμανσης Co,
                                     Ορισμός 1.11)

keep if var1!=.&var2!=.&var3!=.   (Υπολογισμός πίνακα συνδιακύμανσης
correlate var1-var3,cov              με μέθοδο παρατηρήσεων με
matrix H = r(C)                      ελλιπή δεδομένα, Ενότητα 2.1)
```

```
matrix H12 = H[1,2..3]
matrix H22 = H[2..3,2..3]
matrix bhat = H12*invsym(H22)
matrix Bhat=Bhat+bhat

file open B1 using "C:\data\stata\Bhat.txt", write append

   forvalues q=1(1)1{
          forvalues u=1(1)2{
             file write B1 %8z (Bhat[`q',`u'])
                     file write B1 _tab   (Αλλαγή στήλης)
                  }
                 file write B1 _newline   (Αλλαγή γραμμής)
        }
   file close B1

matrix Bhatb2=hadamard(bhat, bhat)
matrix BhatEb2=BhatEb2+Bhatb2


if (v[1,1]> =0 & v[1,2]>=0 & v[1,3]>=0){  (Έλεγχος τιμής ιδιοτιμών)
      matrix Bp1 = Co[1,2..3]              (Διαμέριση πίνακα
      matrix Bp2 = Co[2..3,2..3]            συνδιακύμανσης Co)
      matrix Bposi = Bp1*invsym(Bp2)       (Υπολογισμός συντελεστών
                                            γραμμικής παλινδρόμησης,
                                            Ενότητα 1.5)
      scalar pos = pos+1
      matrix Bpos=Bpos+Bposi
     mat list Bposi

      matrix Bposb2=hadamard(Bposi, Bposi)
            matrix BposEb2=BposEb2+Bposb2
}


if (pos==b){
```

```
     scalar a=2                    (Έξοδος από το while)
}

}

matrix Bfhat=Bhat/m                (Υπολογισμός μέσης τιμής συντελεστών
                                    γραμμικής παλινδρόμησης με χρήση
                                    δεδομένων που προέκυψαν από τη
                                    μέθοδο παρατηρήσεων ελλιπή
                                    δεδομένα, Ορισμός 1.15)

matrix Bmesihat2=BhatEb2/m                  (Υπολογισμός διασποράς
matrix Bfhat2=hadamard(Bfhat, Bfhat)         συντελεστών γραμμικής
matrix Varhat=Bmesihat2-Bfhat2               παλινδρόμησης με χρήση
                                              δεδομένων που προέκυψαν
matrix Bfinalpos=Bpos/pos                     από τη μέθοδο
                                              παρατηρήσεων με ελλιπή
                                              δεδομένα, Ορισμός 1.16)
matrix Bmesipos2=BposEb2/pos
matrix Bfinpos2=hadamard(Bfinalpos, Bfinalpos)
matrix Varpos=Bmesipos2-Bfinpos2

mat list bfull             (Εκτύπωση των μέσων τιμών των
mat list Bfhat              συντελεστών γραμμικής
mat list Bfinalpos         παλινδρόμησης)


mat list Varhat             (Εκτύπωση των διασπορών των
mat list Varpos              συντελεστών γραμμικής
                             παλινδρόμησης)


di pos                      (Εκτύπωση μεταβλητών)
di m
```

ΠΑΡΑΡΤΗΜΑ Β

Στη συνέχεια παρατίθεται σε Stata ο κώδικας που αναπτύχθηκε στο πλαίσιο της στην παρούσας πτυχιακής εργασίας για την υλοποίηση μεθόδων διαταραχής ιδιοτιμών ενός αόριστου πίνακα συνδιακύμανσης, όταν αυτός προκύπτει κατά την εφαρμογή της μεθόδου διαγραφής ανά ζεύγη (βλέπε Σχόλιο 2.3 (ii)). Σύμφωνα με τη θεωρία που αναπτύχθηκε στο Κεφάλαιο 3 οι διορθωτικές μέθοδοι διαταραχής οι οποίες υλοποιούνται είναι οι:

i. ειδικής φασματικής ανάλυσης,
ii. μεταφοράς και
iii. μέσω πίνακα συσχέτισης.

Επίσης, συγκρίνονται οι πίνακες συνδιακύμανσης που προκύπτουν από τις διορθωτικές μεθόδους με κριτήρια απόστασης, νορμών κ.ά.

```
set more off

use "C:\data\stata\cancer_breast.dta", clear

correlate var1-var3,cov
matrix Cf = r(C)
mat list Cf

matrix Bneg=(0,0)            (Αρχικοποίηση πινάκων)
matrix Bnegmet=(0,0)
matrix Bnegcorrelm=(0,0)
matrix BnegEb2=(0,0)
matrix BnegmetEb2=(0,0)
matrix BcorEb2=(0,0)

scalar neg=0                 (Αρχικοποίηση μεταβλητών)
scalar a=1
scalar b=3
scalar k=0
scalar n=15000               (Μέγιστη τιμή επιθυμητών
                              επαναλήψεων)
```

```
while(a==1){
use "C:\data\stata\cancer_breast.dta",clear
local nmiss=556

scalar k=k+1            (Δήλωση μετρητή που μετρά το πλήθος των
                         επαναλήψεων του προγράμματος)

if(k==n){
scalar a=3                           (Έξοδος από το βρόγχο του while)
}

 forvalues j=1(1)`nmiss' {        (Δημιουργία ελλιπών στοιχείων στο
         local z=uniform()         σύνολο δεδομένων)
         local y=`z' *_N
         local y=ceil(`y')
         local z=uniform()
         local x=`z' *3
         local x=ceil(`x')
         replace var`x'=. in `y'
 }

forvalues x=1/3 {              (Δημιουργία πίνακα συνδιακύμανσης
     forvalues y=1/3{           με μέθοδο διαγραφής ανά ζεύγη, Ενότητα 2.3 )
          correlate var`x' var`y',cov
           local var`x'`y'= r(cov_12)
     }
}

 matrix Co =(`var11',`var12', `var13' \ `var21',`var22',
`var23' \ `var31',`var32' ,`var33' )

matrix symeigen X v = Co    (Υπολογισμός ιδιοδιανυσμάτων και ιδιοτιμών
                             πίνακα συνδιακύμανσης Co, Ορισμός 1.11)
mat list Co
mat list v
```

```
matrix SV=(sqrt(`var11'),0,0\0,sqrt(`var22'), 0 \ 0, 0,
sqrt(`var33'))              (Δήλωση πίνακα τυπικής απόκλισης, Πρόταση
                             1.10)

matrix R= inv(SV)*Co*inv(SV)      (Υπολογισμός πίνακα συσχέτισης R,
                                   Πόρισμα 1.1)

matrix symeigen Y r = R        (Υπολογισμός ιδιοδιανυσμάτων και
mat list R                      ιδιοτιμών πίνακα συνδιακύμανσης R,
                                Ορισμός 1.11)

if (v[1,3]<0 & v[1,2]>0){       (Έλεγχος τιμής ιδιοτιμών)

matrix Cnew1=v[1,1]*X[1..3,1]*X[1..3,1]'+
v[1,2]*X[1..3,2]*X[1..3,2]'       (Υλοποίηση μεθόδου μερικής
                                   φασματικής ανάλυσης, Ενότητα 2.1)
matrix Bn1 = Cnew1[1,2..3]
matrix Bn2 = Cnew1[2..3,2..3]
matrix Bnega2 = Bn1*invsym(Bn2)
scalar neg = neg+1
matrix Bneg=Bneg+Bnega2
mat list Bnega2

file open B2 using "C:\data\stata\Bneg.txt", write append

   forvalues c=1(1)1{
     forvalues d=1(1)2{
         file write B2 %8z (Bneg[`c',`d'])
                                                   (Τυπώνει όλα τα Bneg)
              file write B2 _tab                   (Αλλαγή στήλης)
              }
                file write B2 _newline             (Αλλαγή γραμμής)
         }
file close B2

mat list Cnew1
matrix Rnew1= inv(SV)*Cnew1*inv(SV)
```

```
mat list Rnew1

matrix Bnegb2=hadamard(Bnega2, Bnega2)
matrix BnegEb2=BnegEb2+Bnegb2

matrix FFCnew1=Cnew1-Cf
mat list FFCnew1

scalar l=abs(v[1,3])+0.0000001     (Υλοποίηση  μεθόδου μεταφοράς,
scalar w=0.0000001                   Ενότητα 2.2)
matrix L=Co+l*I(3)
matrix L12=L[1,2..3]
matrix L22=L[2..3,2..3]
matrix bm=L12*invsym(L22)
matrix Bnegmet=Bnegmet+bm
mat list bm

file open B3 using "C:\data\stata\Bnegmet.txt", write append
   forvalues e=1(1)1{
       forvalues f=1(1)2{
            file write B3 %8z (Bnegmet[`e',`f'])
                    file write B3 _tab
                }
                file write B3 _newline
           }
file close B3

mat list L
matrix RL= inv(SV)*L*inv(SV)
mat list RL

matrix Bnegmetb2=hadamard(bm, bm)
matrix BnegmetEb2=BnegmetEb2+Bnegmetb2

matrix FFL=L-Cf
mat list FFL
```

```
matrix D1=(r[1,1],0,0\ 0,r[1,2],0\0,0, 0.0000001) (Υλοποίηση
matrix BB=Y*D1*Y'                                  μεθόδου μέσω
 matrix T=(1/sqrt(BB[1,1]) \ 1/sqrt(BB[2,2]) \      πίνακα
1/sqrt(BB[3,3]))                                   συσχέτισης,
matrix TT=T*T'                                     Ενότητα 2.3)
matrix Rnew=hadamard(BB,TT)
matrix symeigen Yr rnew = Rnew
matrix CovRnew=SV*Rnew*SV
matrix CR12=CovRnew[1,2..3]
matrix CR22=CovRnew[2..3,2..3]
matrix bcorrelm=CR12*invsym(CR22)
matrix Bnegcorrelm=Bnegcorrelm+bcorrelm

file open B4 using "C:\data\stata\Bnegcorrelm.txt", write
append
   forvalues g=1(1)1{
     forvalues h=1(1)2{
          file write B4 %8z(Bnegcorrelm[`g',`h'])
                 file write B4 _tab
              }
                file write B4 _newline
           }

file close B4

mat list Rnew
mat list CovRnew

matrix Bcorb2=hadamard(bcorrelm, bcorrelm)
matrix BcorEb2=BcorEb2+Bcorb2

matrix FFCovRnew=CovRnew-Cf
mat list FFCovRnew
}

if (neg==b){
```

```
scalar a=3
}
}
```

(Έξοδος από το βρόγχο του while)

```
matrix Bfineg=Bneg/neg
matrix Bfinegmet=Bnegmet/neg
matrix Bfinegcorrel=Bnegcorrelm/neg
```

(Υπολογισμός μέσης τιμής συντελεστών γραμμικής παλινδρόμησης με χρήση δεδομένων που προέκυψαν από τις μεθόδους ειδικής φασματικής ανάλυσης, μεταφοράς και μέσω πίνακα συσχέτισης, Ορισμός 1.15)

```
matrix Bmesineg2=BnegEb2/neg
matrix Bfineg2=hadamard(Bfineg,Bfineg)
matrix Varneg=Bmesineg2-Bfineg2

matrix Bmesinegmet2=BnegmetEb2/neg
matrix Bfinegmet2=hadamard(Bfinegmet,Bfinegmet)
matrix Varnegmet=Bmesinegmet2-Bfinegmet2
```

(Υπολογισμός διασποράς συντελεστών γραμμικής παλινδρόμησης με χρήση δεδομένων που προέκυψαν από τις μεθόδους ειδικής φασματικής ανάλυσης, μεταφοράς και μέσω πίνακα συσχέτισης, Ορισμός 1.16)

```
matrix Bmesicor2=BcorEb2/neg
matrix Bfincor2=hadamard(Bfinegcorrel, Bfinegcorrel)
matrix Varnegcorrel =Bmesicor2-Bfincor2
```

(Υλοποίηση κριτηρίου διαφοράς με χρήση απόλυτων τιμών, Ενότητα 3.2.2)

```
matrix
MSADChat=(abs(FFChat[1,1]+FFChat[2,1]+FFChat[3,1]+FFChat[2,2]+
FFChat[3,2]+FFChat[3,3])/6)
matrix
MSADRhat=(abs(FFRhat[1,1]+FFRhat[2,1]+FFRhat[3,1]+FFRhat[2,2]+
FFRhat[3,2]+FFRhat[3,3])/6)
```

```
matrix
MSADCnew1=(abs(FFCnew1[1,1]+FFCnew1[2,1]+FFCnew1[3,1]+FFCnew1[
2,2]+FFCnew1[3,2]+FFCnew1[3,3])/6)
matrix
MSADRnew1=(abs(FFRnew1[1,1]+FFRnew1[2,1]+FFRnew1[3,1]+FFRnew1[
2,2]+FFRnew1[3,2]+FFRnew1[3,3])/6)
matrix
MSADL=(abs(FFL[1,1]+FFL[2,1]+FFL[3,1]+FFL[2,2]+FFL[3,2]+FFL[3,
3])/6)
matrix
MSADRL=(abs(FFRL[1,1]+FFRL[2,1]+FFRL[3,1]+FFRL[2,2]+FFRL[3,2]+
FFRL[3,3])/6)
matrix
MSADCovRnew=(abs(FFCovRnew[1,1]+FFCovRnew[2,1]+FFCovRnew[3,1]+
FFCovRnew[2,2]+FFCovRnew[3,2]+FFCovRnew[3,3])/6)
matrix
MSADRnew=(abs(FFRnew[1,1]+FFRnew[2,1]+FFRnew[3,1]+FFRnew[2,2]+
FFRnew[3,2]+FFRnew[3,3])/6)

mat list MSADChat
mat list MSADRhat
mat list MSADCnew1
mat list MSADRnew1
mat list MSADL
mat list MSADRL
mat list MSADCovRnew
mat list MSADRnew

mat list Bfineg         (Εκτύπωση των μέσων τιμών των συντελεστών
mat list Bfinegmet        γραμμικής παλινδρόμησης)
mat list Bfinegcorrel

mat list Varneg         (Εκτύπωση των διασπορών των συντελεστών
mat list Varnegmet        γραμμικής παλινδρόμησης)
mat list Varnegcorrel

di neg                  (Εκτύπωση μεταβλητών)
di w
```

```
di k

mata                    (Είσοδος στο Mata)

FFCnew1=st_matrix("FFCnew1")      (Δήλωση των πινάκων που θα
FFL=st_matrix("FFL")               χρησιμοποιηθούν από το stata)
FFCovRnew=st_matrix("FFCovRnew")

FFCnew1                 (Εκτύπωση των πινάκων στην οθόνη)
FFL
FFCovRnew

normCnew1=norm(FFCnew1,0)     (Εύρεση της νόρμας Frobenius των πινάκων
normL=norm(FFL,0)              που δηλώθηκαν παραπάνω, Ενότητα 3.2.1)
normCovRnew=norm(FFCovRnew,0)

normCnew1               (Εκτύπωση των νορμών στην οθόνη)
normL
normCovRnew

MSDCnew1=normCnew1/9
MSDRnew1=normRnew1/9
MSDL=normL/9
MSDRL=normRL/9
MSDCovRnew=normCovRnew/9
MSDRnew=normRnew/9

MSDCnew1
MSDRnew1
MSDL
MSDRL
MSDCovRnew
MSDRnew

st_matrix("nCnew1",normCnew1)  (Τα αποτελέσματα περνάν στο stata)
st_matrix("nRnew1",normRnew1)
```

```
st_matrix("nL",normL)
st_matrix("nRL",normRL)

st_matrix("nCovRnew",normCovRnew)
st_matrix("nRnew",normRnew)

end                         (Έξοδος από το mata)
```

ΠΑΡΑΡΤΗΜΑ Γ

Στο παρόν παράρτημα παρατίθενται δυο σύνολα δεδομένων, (http://archive.ics.uci.edu/ml/) [28], που τιτλοφορούνται ως ακολούθως:

i. ηπατικές διαταραχές και

ii. καρκίνος του μαστού.

Γ.1 Ηπατικές διαταραχές

Ένα από τα σύνολα δεδομένων που χρησιμοποιήθηκε στην παρούσα πτυχιακή εργασία παρατίθεται στη συνέχεια. Πρόκειται για ένα σύνολο δεδομένων το οποίο αποτελείται από 345 παρατηρήσεις και 3 μεταβλητές (var1, var2, var3) και παρουσιάζει στοιχεία σχετικά με τις ηπατικές διαταραχές (liver disorders).

Για την ακρίβεια οι μεταβλητές var1, var2, var3 συνιστούν δεδομένα που έχουν προέλθει από εξετάσεις αίματος σχετικών με τις ηπατικές διαταραχές και κάθε παρατήρηση αναφέρεται σε διαφορετικό άτομο. Η μεταβλητή var1 περιγράφει την αμινιτρανσφεράση της αλανίνης (alanine aminotransferase), η var2 την αμιινοτρανσφεράση του ασπαρτικού (aspartate aminotransferase) και η var3 την τρανσπεπτιδάση του γάμμα γλουταμινικού (gamma-glutamyl transpeptidase). Αξίζει να σημειωθεί ότι το pvalue των συγκεκριμένων μεταβλητών είναι ίσο με την τιμή 0.0000000.

Τέλος, θα συνιστούσε παράλειψη να μην αναφερθεί η πηγή του συνόλου δεδομένων. Πρόκειται για τη βάση UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/Liver+Disorders) [29]. Επίσης, το πλήθος των μεταβλητών του συνόλου δεδομένων που βρέθηκε στη βάση UCI αριθμούνταν σε 7 αλλά επιλέχθηκαν 3 μεταβλητές εξ’ αυτών κατά την υλοποίηση του κώδικα.

	var1	var2	var3
1	45	27	31
2	59	32	23
3	33	16	54
4	34	24	36
5	12	28	10
6	13	17	17
7	20	17	9
8	21	11	11
9	22	20	7
10	25	19	5
11	13	24	15
12	17	17	15
13	61	32	13
14	25	19	18
15	29	20	11
16	20	31	18
17	23	16	10
18	17	17	16
19	20	20	56
20	11	33	11
21	35	13	19
22	35	20	20
23	23	27	5
24	18	19	19
25	47	33	97
26	24	13	26
27	28	15	18
28	20	21	10
29	17	13	19
30	31	19	16
31	28	16	17
32	31	23	42
33	32	18	29
34	28	21	40
35	22	18	11
36	155	68	82
37	47	33	22
38	14	19	9
39	25	26	30
40	24	20	38
41	68	37	44
42	26	39	42
43	18	25	13
44	18	14	16
45	25	14	18
46	19	21	13
47	14	16	10
48	29	25	50
49	24	22	11
50	29	25	38
51	19	23	16
52	24	22	11
53	64	36	90
54	11	23	43
55	21	19	30
56	23	29	15
57	21	21	19
58	20	22	19
59	26	26	36
60	24	24	24
61	33	34	27
62	14	15	18
63	4	8	13
64	15	22	11
65	32	19	27
66	26	20	19
67	5	17	14
68	26	24	23
69	19	30	13
70	16	21	14
71	29	23	57
72	35	19	35
73	27	22	9
74	34	21	22
75	20	21	16
76	19	14	42
77	43	28	156
78	28	19	30
79	9	25	16
80	27	25	30

81	34	30	64
82	41	22	48
83	42	33	16
84	21	18	26
85	30	27	297
86	21	18	26
87	33	26	29
88	17	25	9
89	22	20	11
90	20	25	7
91	27	21	26
92	20	20	6
93	45	32	30
94	23	12	15
95	42	30	20
96	23	17	27
97	32	37	53
98	23	18	104
99	26	18	36
100	26	26	24
101	25	20	25
102	18	13	81
103	22	27	10
104	18	23	13
105	25	19	14
106	27	29	17
107	37	23	27
108	33	15	18
109	17	5	7
110	29	20	50
111	22	55	9
112	28	23	21
113	26	23	17
114	24	25	17
115	34	49	169
116	11	15	8
117	19	20	21
118	17	21	14
119	15	17	25
120	15	16	16
121	48	39	42
122	17	27	23
123	14	20	14
124	14	21	24
125	29	27	31
126	10	16	16
127	24	21	48
128	28	28	82
129	19	21	13
130	19	14	22
131	21	16	33
132	26	18	18
133	47	39	107
134	113	45	150
135	15	19	12
136	21	11	15
137	16	20	24
138	21	23	22
139	25	23	112
140	17	19	11
141	10	26	20
142	26	22	28
143	25	26	15
144	37	21	36
145	13	14	15
146	37	19	50
147	70	26	36
148	64	42	76
149	20	23	20
150	25	26	15
151	67	38	92
152	27	24	37
153	12	26	21
154	9	15	16
155	29	23	76
156	35	23	69
157	87	50	67
158	57	25	44
159	36	34	48
160	17	19	19
161	21	20	68
162	18	25	17
163	20	26	33
164	31	28	41

165	19	17	14
166	24	29	29
167	70	32	84
168	58	35	120
169	58	47	62
170	29	22	52
171	33	19	26
172	42	23	73
173	22	21	21
174	18	17	6
175	77	39	114
176	29	22	52
177	25	25	66
178	17	20	14
179	85	48	200
180	26	22	53
181	59	47	60
182	22	28	123
183	86	41	31
184	22	24	26
185	31	34	73
186	50	64	55
187	53	33	94
188	22	28	49
189	39	37	108
190	81	34	201
191	25	21	14
192	23	18	12
193	41	20	32
194	27	20	15
195	17	13	5
196	25	20	12
197	27	15	12
198	31	25	15
199	30	14	21
200	26	18	8
201	19	19	22
202	27	15	12
203	40	26	56
204	37	28	14
205	35	22	135
206	23	14	35

207	24	23	11
208	19	25	19
209	33	32	18
210	17	8	9
211	30	26	17
212	24	18	31
213	34	21	27
214	17	27	7
215	23	15	12
216	27	34	14
217	11	12	10
218	41	20	53
219	29	22	18
220	35	29	42
221	28	25	35
222	18	15	24
223	22	26	11
224	19	20	14
225	12	17	14
226	18	24	16
227	41	27	36
228	52	29	62
229	38	28	48
230	36	27	59
231	30	18	27
232	30	30	22
233	148	75	159
234	20	25	38
235	17	20	71
236	25	21	33
237	18	21	22
238	25	20	31
239	14	16	13
240	16	23	23
241	15	13	22
242	13	20	13
243	26	30	13
244	33	27	34
245	32	26	13
246	21	19	14
247	23	15	19
248	32	14	8

249	28	22	10
250	45	24	85
251	21	22	37
252	31	19	115
253	24	23	14
254	41	30	48
255	43	32	38
256	33	20	22
257	24	15	18
258	28	24	21
259	31	17	17
260	28	29	31
261	70	24	64
262	18	17	26
263	33	21	28
264	18	18	23
265	37	19	70
266	30	26	25
267	14	15	10
268	26	25	27
269	37	27	34
270	20	23	18
271	9	15	16
272	38	27	17
273	12	22	11
274	20	20	16
275	21	17	26
276	35	26	33
277	31	20	84
278	83	49	51
279	25	22	35
280	27	25	53
281	29	21	26
282	32	22	14
283	20	22	9
284	17	19	23
285	21	26	21
286	77	56	52
287	11	14	21
288	27	21	29
289	46	32	39
290	17	17	46

291	31	18	37
292	19	13	16
293	14	20	19
294	29	35	108
295	52	47	52
296	15	33	55
297	22	23	34
298	33	27	55
299	23	18	19
300	154	82	121
301	38	26	23
302	10	17	12
303	20	16	12
304	14	21	49
305	47	23	37
306	25	26	31
307	33	34	115
308	41	29	23
309	15	14	14
310	31	26	32
311	51	33	92
312	43	43	82
313	11	16	54
314	24	18	19
315	38	19	15
316	58	42	203
317	103	57	114
318	10	26	20
319	27	25	43
320	21	27	47
321	16	22	28
322	30	24	39
323	34	78	203
324	12	18	14
325	33	23	12
326	76	32	24
327	46	30	33
328	36	19	25
329	28	24	31
330	18	23	44
331	20	25	225
332	17	20	53

333	25	24	28
334	56	35	126
335	45	21	48
336	53	43	203
337	37	22	37
338	20	25	23
339	35	34	37
340	15	23	11
341	52	43	55
342	26	24	41

Γ.2 Καρκίνος του μαστού

Πρόκειται για ένα σύνολο δεδομένων το οποίο αποτελείται από 569 παρατηρήσεις και 6 μεταβλητές και παρουσιάζει στοιχεία σχετικά με τον καρκίνο του μαστού (breast cancer).

Για την ακρίβεια οι μεταβλητές var1, var2, var3, var4, var5 και var6 είναι χαρακτηριστικά που υπολογίζονται από μια ψηφιακή εικόνα μιας μάζας μαστού η οποία έχει υποστεί βιοψία δια λεπτής βελόνης (FNA). Αυτές οι μεταβλητές περιγράφουν χαρακτηριστικά των πυρήνων των κυττάρων που υπάρχουν στην εικόνα. Ειδικότερα, η μεταβλητή var1 περιγράφει το πόσο λείος είναι, η var2 το πόσο συμπαγές είναι $(περί\,μετρος^2 / περιοχή - 1.0)$, η var3 το πόσο κοίλος είναι, η var4 την ακτίνα (μέσος όρος των αποστάσεων από το κέντρο στα σημεία της περιμέτρου), η var5 το τυπικό σφάλμα της ακτίνας και η var6 το τυπικό σφάλμα της συμμετρίας.

Οι μεταβλητές χρησιμοποιήθηκαν σε δυο ζεύγη των τριών μεταβλητών. Το πρώτο ζεύγος αποτελείται από τις μεταβλητές var1, var2 και var3 των οποίων το pvalue είναι ίσο με την τιμή 0.00000000 και το άλλο ζεύγος αποτελείται από τις var4, var5 και var6 των οποίων το pvalue έχει την τιμή 0.808.

Τέλος, αξίζει να σημειωθεί ότι η πηγή του συγκεκριμένου συνόλου δεδομένων είναι η βάση UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)) [30] καθώς επίσης πως το πλήθος των μεταβλητών του συνόλου δεδομένων που βρέθηκε στη βάση UCI αριθμούνταν σε 32 αλλά επιλέχθηκαν 6 μεταβλητές εξ' αυτών κατά την υλοποίηση του κώδικα.

	var1	var2	var3	var4	var5	var6
1	.1184	.2776	.3001	17.99	8.589	25.38
2	.08474	.07864	.0869	20.57	3.398	24.99
3	.1096	.1599	.1974	19.69	4.585	23.57
4	.1425	.2839	.2414	11.42	3.445	14.91
5	.1003	.1328	.198	20.29	5.438	22.54
6	.1278	.17	.1578	12.45	2.217	15.47
7	.09463	.109	.1127	18.25	3.18	22.88
8	.1189	.1645	.09366	13.71	3.856	17.06
9	.1273	.1932	.1859	13	2.406	15.49
10	.1186	.2396	.2273	12.46	2.039	15.09
11	.08206	.06669	.03299	16.02	2.466	19.19
12	.0971	.1292	.09954	15.78	3.564	20.42
13	.0974	.2458	.2065	19.17	11.07	20.96
14	.08401	.1002	.09938	15.85	2.903	16.84
15	.1131	.2293	.2128	13.73	2.061	15.03
16	.1139	.1595	.1639	14.54	2.879	17.46
17	.09867	.072	.07395	14.68	3.195	19.07
18	.117	.2022	.1722	16.13	3.854	20.96
19	.09831	.1027	.1479	19.81	5.865	27.32
20	.09779	.08129	.06664	13.54	2.058	15.11
21	.1075	.127	.04568	13.08	1.383	14.5
22	.1024	.06492	.02956	9.504	1.909	10.23
23	.1073	.2135	.2077	15.34	3.384	18.07
24	.09428	.1022	.1097	21.16	4.303	29.17
25	.1121	.1457	.1525	16.65	5.455	26.46
26	.1186	.2276	.2229	17.14	7.276	22.25
27	.1054	.1868	.1425	14.58	2.11	17.62
28	.0944	.1066	.149	18.61	5.632	21.31
29	.1082	.1697	.1683	15.3	3.498	20.27
30	.09847	.1157	.09875	17.57	4.655	20.01
31	.1064	.1887	.2319	18.63	5.574	23.15
32	.1109	.1516	.1218	11.84	3.475	16.82
33	.1197	.1496	.2417	17.02	3.999	20.88
34	.09401	.1719	.1657	19.27	3.528	24.15
35	.104	.1559	.1354	16.13	2.183	20.21
36	.0961	.1336	.1348	16.74	3.008	20.01
37	.09823	.1098	.1319	14.25	2.657	15.89
38	.08983	.03766	.02562	13.03	1.17	13.3
39	.09387	.05131	.02398	14.99	8.077	14.99
40	.1016	.1255	.1063	13.48	1.545	15.53
41	.08162	.06031	.0311	13.44	1.572	15.93

42	.1227	.1218	.1044	10.95	1.822	12.84
43	.09081	.219	.2107	19.07	8.83	24.09
44	.1041	.1436	.09847	13.28	2.427	17.38
45	.09714	.1047	.08259	13.17	1.334	16.23
46	.1099	.1686	.1974	18.65	4.293	22.82
47	.086	.05943	.01588	8.196	1.094	8.964
48	.1158	.1231	.1226	13.17	1.897	15.67
49	.1031	.09092	.06592	12.05	1.848	13.76
50	.08752	.07698	.04751	13.49	1.735	15.15
51	.08637	.04966	.01657	11.76	2.635	12.98
52	.07685	.06059	.01857	13.64	1.449	14.67
53	.08261	.04751	.01972	11.94	1.52	13.1
54	.1148	.1485	.1772	18.22	4.877	20.6
55	.09056	.07081	.05253	15.1	2.097	18.1
56	.09524	.05473	.03036	11.52	2.183	12.84
57	.1053	.1267	.1323	19.21	4.837	26.14
58	.1137	.1365	.1293	14.71	2.735	17.87
59	.0806	.03789	.000692	13.05	2.595	14.23
60	.09752	.05272	.02061	8.618	1.046	9.507
61	.1134	.08061	.01084	10.17	3.312	11.02
62	.1243	.08963	.03	8.598	2.493	9.565
63	.1049	.2008	.2135	14.25	5.373	17.67
64	.07721	.08751	.05988	9.173	2.608	10.01
65	.1122	.1262	.1128	12.68	2.927	17.09
66	.1172	.1479	.1267	14.78	2.45	17.31
67	.1044	.07773	.02172	9.465	1.66	10.41
68	.08139	.04701	.03709	11.31	1.831	12.33
69	.1066	.1413	.313	9.029	1.885	10.31
70	.09831	.05234	.03653	12.78	1.471	13.46
71	.09009	.1029	.108	18.94	5.486	24.86
72	.09783	.1531	.08606	8.888	3.168	9.733
73	.1071	.183	.1692	17.2	3.705	23.32
74	.1007	.128	.07789	13.8	1.957	16.57
75	.09172	.06829	.03372	12.31	1.74	14.11
76	.09168	.08424	.09769	16.07	5.029	19.77
77	.1291	.1047	.06877	13.53	2.652	14.08
78	.1065	.2146	.1684	18.05	6.311	22.39
79	.1286	.3454	.3754	20.18	8.649	23.37
80	.09934	.09546	.03889	12.86	1.778	14.24
81	.1102	.09362	.04591	11.45	2.077	13.11
82	.1078	.1535	.1169	13.34	1.535	15.53
83	.1063	.2665	.3339	25.22	7.382	30

84	.1215	.1791	.1937	19.1	5.801	20.33
85	.09723	.07165	.04151	12	1.441	13.67
86	.09874	.1053	.1335	18.46	4.782	22.93
87	.09444	.09947	.1204	14.48	3.301	16.21
88	.09029	.1206	.1468	19.02	3.055	24.56
89	.08772	.09445	.06015	12.36	2.203	13.83
90	.1132	.1339	.09966	14.64	3.814	16.34
91	.08974	.08606	.03102	14.62	2.279	16.11
92	.092	.1036	.1122	15.37	2.075	16.43
93	.07355	.05055	.03261	13.27	2.701	16.36
94	.1022	.08165	.03974	13.45	2.099	15.1
95	.1039	.1553	.17	15.06	3.706	18.23
96	.09078	.1313	.1465	20.26	4.554	24.22
97	.1045	.07057	.0249	12.18	2.41	12.83
98	.1024	.05301	.006829	9.787	2.132	10.92
99	.08983	.07525	.04196	11.6	1.475	13.06
100	.09752	.1141	.09388	14.42	2.376	16.33
101	.09488	.08511	.08625	13.61	2.861	16.99
102	.117	.07568	0	6.981	1.553	7.93
103	.08013	.04038	.02383	12.18	1.183	13.34
104	.1005	.09697	.06154	9.876	1.528	10.76
105	.09989	.08578	.02995	10.49	2.302	11.54
106	.1398	.1765	.2071	13.11	2.41	16.31
107	.1142	.1017	.0707	11.64	2.155	13.14
108	.08477	.06815	.02643	12.36	.8484	13.29
109	.1326	.2768	.4264	22.27	10.05	28.4
110	.08759	.06575	.05133	11.34	1.597	13.01
111	.1037	.08404	.04334	9.777	2.747	11.05
112	.09933	.1209	.1065	12.63	2.711	13.33
113	.07837	.2233	.3003	14.26	3.399	15.3
114	.1122	.1303	.06476	10.51	2.041	11.16
115	.115	.08201	.04132	8.726	1.354	9.628
116	.09768	.07849	.03328	11.93	2	13.67
117	.09462	.1243	.09263	8.95	3.28	9.414
118	.1162	.1649	.169	14.87	2.989	18.81
119	.1155	.1752	.2133	15.78	3.598	20.19
120	.08402	.06722	.07293	17.95	3.357	20.58
121	.09373	.06685	.03512	11.41	1.103	12.82
122	.1054	.11	.1457	18.66	4.895	22.25
123	.1447	.2867	.4268	24.25	9.807	26.02
124	.1101	.1099	.08842	14.5	1.928	15.7
125	.07115	.07325	.08092	13.37	1.223	14.26

126	.08785	.06136	.0142	13.85	1.495	15.49
127	.09258	.07862	.05285	13.61	1.752	16.89
128	.08217	.08028	.09271	19	5.216	22.32
129	.115	.1807	.1138	15.1	2.796	16.11
130	.1015	.1589	.2545	19.79	2.765	22.63
131	.1066	.09509	.02855	12.19	1.973	13.34
132	.1092	.1223	.1466	15.46	3.094	19.26
133	.1008	.1284	.1043	16.16	2.844	19.47
134	.09462	.09462	.07135	15.71	1.972	17.5
135	.0943	.09709	.1153	18.45	3.766	22.52
136	.09055	.05761	.04711	12.77	1.457	14.49
137	.1051	.06095	.03592	11.71	3.258	13.33
138	.09639	.06889	.03503	11.43	1.143	12.32
139	.1167	.1305	.1539	14.95	8.419	18.55
140	.1164	.1136	.04635	11.28	1.851	11.92
141	.0925	.04102	0	9.738	1.218	10.62
142	.09721	.1137	.09447	16.11	4.533	19.92
143	.1092	.09486	.02031	11.43	1.937	12.78
144	.08677	.09509	.04894	12.9	1.689	14.48
145	.07793	.05139	.02251	10.75	1.806	11.95
146	.1152	.1296	.0371	11.9	3.021	13.15
147	.1091	.17	.1659	11.8	2.281	13.74
148	.08138	.1167	.0905	14.95	3.271	16.25
149	.0997	.1021	.08487	14.44	2.12	15.85
150	.07944	.06376	.02881	13.74	1.573	15.34
151	.1135	.07589	.03136	13	2.873	14.16
152	.09405	.1305	.1321	8.219	1.243	9.092
153	.1072	.1599	.4108	9.731	4.073	11.02
154	.09754	.05113	.01982	11.15	1.429	11.99
155	.09384	.08498	.09293	13.15	1.819	14.77
156	.08654	.06679	.03885	12.25	1.484	13.59
157	.1115	.1665	.1855	17.68	5.54	20.47
158	.07445	.07223	.0515	16.84	3.479	18.22
159	.09311	.05241	.01972	12.06	1.171	13.14
160	.07515	.03718	.00309	10.9	1.808	12.36
161	.1089	.1141	.06843	11.75	3.926	13.32
162	.08694	.1185	.1193	19.19	6.971	22.03
163	.112	.1666	.2508	19.59	4.792	26.73
164	.1012	.1015	.0537	12.34	1.955	13.58
165	.08439	.1145	.1324	23.27	4.603	28.01
166	.08421	.05352	.01947	14.97	1.286	15.98
167	.09594	.05736	.02531	10.8	1.126	11.6

168	.08865	.09182	.08422	16.78	4.129	20.05
169	.1049	.1603	.2159	17.47	7.337	23.14
170	.09855	.07885	.02602	14.97	1.893	16.11
171	.1028	.06981	.03987	12.32	1.67	13.5
172	.09048	.06288	.05858	13.43	3.142	17.98
173	.1257	.1555	.2032	15.46	2.805	18.79
174	.1006	.05743	.02363	11.08	1.377	11.35
175	.08792	.04302	0	10.66	2.155	11.54
176	.09138	.04276	0	8.671	1.435	9.262
177	.09699	.1294	.1307	9.904	3.132	11.26
178	.09831	.1556	.1793	16.46	2.482	17.79
179	.06251	.01938	.001595	13.01	1.101	14
180	.08739	.03774	.009193	12.81	1.778	13.63
181	.1094	.1914	.2871	27.22	5.82	33.12
182	.1141	.2832	.2487	21.09	4.414	26.68
183	.09597	.08799	.06593	15.7	2.406	20.11
184	.09059	.08155	.06181	11.41	1.902	12.37
185	.09057	.1052	.05375	15.28	1.344	17.8
186	.09267	.04695	.001597	10.08	2.68	11.87
187	.08588	.08468	.08169	18.31	1.817	21.31
188	.09774	.06141	.03809	11.71	1.742	13.01
189	.1007	.05562	.02353	11.81	1.011	12.57
190	.0808	.07253	.03844	12.3	1.687	13.35
191	.1075	.2413	.1981	14.22	2.112	15.74
192	.08749	.06601	.03112	12.77	5.118	13.75
193	.0695	.02344	0	9.72	2.23	9.968
194	.1034	.1353	.1085	12.34	2.642	15.65
195	.1044	.198	.1697	14.86	3.591	16.08
196	.07941	.05366	.03873	12.91	1.493	13.88
197	.12	.1267	.1385	13.77	4.906	16.39
198	.07371	.08642	.1103	18.08	4.312	19.76
199	.08523	.1428	.1114	19.18	3.833	23.36
200	.09872	.1206	.118	14.45	1.446	18.33
201	.09586	.08087	.04187	12.23	2.308	14.44
202	.08968	.1198	.1036	17.54	3.088	20.42
203	.1141	.2084	.3523	23.29	4.667	25.12
204	.1323	.1768	.1558	13.81	3.909	19.2
205	.09965	.1058	.08005	12.47	2.497	14.97
206	.08876	.09588	.0755	15.12	1.974	17.77
207	.1089	.07232	.01756	9.876	1.517	10.42
208	.08772	.07304	.0695	17.01	4.106	19.8
209	.1002	.1483	.08705	13.11	1.491	14.55

210	.08182	.0623	.05892	15.27	1.525	17.38
211	.0909	.1348	.164	20.58	7.029	23.24
212	.08871	.069	.02669	11.84	1.444	13.3
213	.1142	.1516	.3201	28.11	21.98	28.11
214	.1006	.1146	.1682	17.42	3.767	18.07
215	.09463	.1306	.1115	14.19	3.534	16.86
216	.1026	.1517	.09901	13.86	1.933	15.75
217	.09363	.1154	.06636	11.89	2.087	13.25
218	.08054	.05907	.05774	10.2	2.747	11.48
219	.09383	.1306	.1272	19.8	6.487	25.73
220	.0842	.113	.1145	19.53	4.722	27.9
221	.09646	.08711	.03888	13.65	1.391	15.34
222	.1051	.1192	.0786	13.56	2.011	14.98
223	.1061	.08502	.01768	10.18	1.641	11.17
224	.1025	.1204	.1147	15.75	2.244	19.56
225	.08445	.04994	.03554	13.27	2.044	15.14
226	.09906	.07624	.05724	14.34	3.763	16.77
227	.1053	.07722	.006643	10.44	1.208	11.52
228	.08371	.1096	.06505	15	2.276	16.41
229	.07903	.07529	.05438	12.62	1.445	14.2
230	.1088	.1799	.1695	12.83	2.257	15.2
231	.1141	.1572	.191	17.05	2.153	19.59
232	.06883	.03813	.01633	11.32	1.059	12.08
233	.0778	.03574	.004967	11.22	1.489	12.36
234	.09159	.1074	.1554	20.51	3.767	24.47
235	.08464	.04087	.01652	9.567	1.215	10.51
236	.0907	.06945	.01462	14.03	1.667	15.33
237	.09509	.1682	.195	23.21	7.247	31.01
238	.08355	.08348	.09042	20.48	5.144	24.22
239	.08223	.1039	.1103	14.22	2.105	15.75
240	.09812	.1298	.1417	17.46	3.002	22.51
241	.09423	.0663	.04705	13.64	1.996	14.85
242	.07926	.03393	.01053	12.42	.757	13.2
243	.09592	.1325	.1548	11.3	2.369	12.58
244	.08043	.06807	.04697	13.75	2.829	15.01
245	.1027	.1558	.2049	19.4	4.037	21.65
246	.107	.05971	.04831	10.48	2.517	11.48
247	.07215	.04524	.04336	13.2	.873	13.94
248	.0876	.1346	.1374	12.89	2.393	14.39
249	.09657	.07234	.02379	10.65	1.497	12.25
250	.1013	.07808	.04328	11.52	1.686	12.65
251	.1007	.1606	.2712	20.94	6.372	25.58

252	.09345	.05991	.02638	11.5	2.684	12.97
253	.1062	.1849	.2417	19.73	4.115	25.28
254	.1008	.1041	.1266	17.3	2.193	19.85
255	.1035	.1188	.1379	19.45	3.797	25.7
256	.1096	.1279	.09789	13.96	2.563	16.39
257	.0926	.2063	.1784	19.55	7.158	25.05
258	.1335	.2284	.2448	15.32	4.061	17.73
259	.1109	.3114	.3176	15.66	10.12	19.85
260	.1063	.1639	.1751	15.53	1.903	18.49
261	.1	.1088	.1519	20.31	2.587	24.33
262	.08662	.0629	.02891	17.35	2.577	19.85
263	.08999	.1273	.09697	17.29	6.146	20.39
264	.0784	.05616	.04209	15.61	1.534	17.91
265	.09726	.08995	.09061	17.19	2.819	21.58
266	.09469	.1143	.1367	20.73	7.749	32.49
267	.09688	.1147	.06387	10.6	3.43	11.88
268	.07956	.08259	.04072	13.59	2.591	14.8
269	.09425	.06219	.039	12.87	1.546	13.9
270	.1082	.1289	.08448	10.71	2.23	11.69
271	.06429	.02675	.00725	14.29	.8439	14.91
272	.09834	.07608	.03265	11.29	1.164	12.32
273	.09401	.1961	.2195	21.75	8.867	28.19
274	.09037	.04689	.01103	9.742	1.75	10.75
275	.08855	.07027	.05699	17.93	2.765	20.92
276	.1225	.0721	.05929	11.89	4.021	12.4
277	.09379	.03872	.001487	11.33	1.565	12.2
278	.08923	.05884	.0802	18.81	2.363	19.96
279	.07948	.04052	.01997	13.59	1.683	15.5
280	.09516	.07688	.04479	13.85	1.83	14.98
281	.102	.1453	.1921	19.16	4.321	23.72
282	.07813	.0434	.02245	11.74	3.717	13.31
283	.1037	.1442	.1626	19.4	2.903	23.79
284	.1066	.1802	.1948	16.24	2.464	18.55
285	.07818	.0958	.1115	12.89	2.347	13.9
286	.08393	.04216	.00186	12.58	1.721	13.5
287	.08605	.1011	.06574	11.94	3.198	13.24
288	.06955	.03729	.0226	12.89	1.115	13.62
289	.0802	.1181	.09274	11.26	2.877	11.86
290	.08713	.05008	.02399	11.37	1.954	12.36
291	.08757	.1676	.1362	14.41	4.36	15.77
292	.08992	.09823	.0594	14.96	2.171	16.25
293	.1005	.07943	.06155	12.95	1.231	13.74

294	.08372	.05642	.02688	11.85	1.234	13.06
295	.09667	.08393	.01288	12.72	1.34	13.5
296	.09198	.06221	.01063	13.77	1.479	14.67
297	.08518	.04721	.01236	10.91	1.267	11.37
298	.09968	.05914	.02685	11.76	4.138	13.36
299	.06576	.0522	.02475	14.26	1.661	16.22
300	.1015	.06797	.02495	10.51	2.289	10.93
301	.115	.1642	.2197	19.53	7.237	25.93
302	.08451	.1014	.0683	12.46	2.579	13.46
303	.108	.1838	.2283	20.09	7.804	23.68
304	.1068	.06678	.02297	10.49	1.035	11.06
305	.08853	.07694	.03344	11.46	2.475	12.68
306	.07474	.05688	.01974	11.6	1.961	12.44
307	.08511	.05251	.001461	13.2	1.204	14.41
308	.07005	.03116	.003681	9	1.144	9.699
309	.07376	.03614	.002758	13.5	1.509	14.97
310	.08352	.03735	.004559	13.05	2.567	14.73
311	.08814	.05253	.01583	11.7	1.109	12.61
312	.07618	.03515	.01447	14.61	1.954	16.46
313	.08794	.07948	.04052	12.76	2.346	14.19
314	.08597	.05969	.01367	11.54	1.107	12.34
315	.1074	.05847	0	8.597	2.222	8.952
316	.08511	.03834	.004473	12.49	1.047	13.34
317	.07734	.03212	.01123	12.18	1.438	12.85
318	.09746	.1117	.113	18.22	2.547	21.84
319	.09968	.1972	.1975	9.042	3.769	10.06
320	.07557	.03454	.01342	12.43	2.487	12.9
321	.1061	.1111	.06726	10.25	1.597	11.28
322	.0802	.08564	.1155	20.16	3.868	23.06
323	.1134	.08834	.038	12.86	1.614	14.04
324	.117	.1875	.2565	20.34	4.012	25.3
325	.08673	.06545	.01994	12.2	1.959	13.75
326	.1028	.07664	.03193	12.67	1.566	13.71
327	.09309	.05306	.01765	14.11	1.558	15.53
328	.07683	.03892	.001546	12.03	1.466	13.07
329	.1169	.1319	.1478	16.27	3.27	19.28
330	.1165	.1283	.1799	16.26	2.961	17.73
331	.09491	.1371	.1204	16.03	2.629	18.76
332	.09579	.1125	.07107	12.98	2.465	14.42
333	.1054	.06779	.005006	11.22	1.959	11.98
334	.08306	.04458	.000974	11.25	1.529	12.76
335	.08313	.04202	.007756	12.3	1.199	13.35

336	.1119	.1056	.1508	17.06	6.076	20.99
337	.09462	.09965	.03738	12.99	.9219	13.72
338	.09116	.1402	.106	18.77	4.369	24.54
339	.1007	.07326	.02511	10.05	1.778	11.16
340	.1069	.1283	.2308	23.51	6.462	30.67
341	.09751	.1139	.08007	14.42	2.677	16.67
342	.08481	.09228	.08422	9.606	1.429	10.75
343	.1033	.09097	.05397	11.06	1.355	11.92
344	.09797	.1339	.1863	19.68	5.173	22.75
345	.115	.07281	.04006	11.71	2.355	13.06
346	.09882	.09159	.03581	10.26	2.394	10.88
347	.08386	.05794	.00751	12.06	1.559	13.64
348	.08875	.0778	.04608	14.76	2.537	17.27
349	.09076	.05886	.02587	11.47	1.09	12.51
350	.1158	.1206	.01171	11.95	2.455	12.81
351	.07561	.0363	.008306	11.66	2.225	13.28
352	.1243	.2364	.2914	15.75	3.477	17.36
353	.1149	.2363	.3368	25.73	7.222	33.13
354	.1024	.09769	.1235	15.08	4.174	18.51
355	.07274	.06064	.04505	11.14	3.33	12.12
356	.0876	.1038	.103	12.56	3.212	13.37
357	.1082	.1304	.09603	13.05	2.59	14.19
358	.08743	.05492	.01502	13.87	1.737	15.11
359	.08293	.07698	.04721	8.878	4.277	9.981
360	.1009	.05956	.0271	9.436	3.267	12.02
361	.07436	.0265	.001194	12.54	2.329	13.72
362	.08582	.06373	.03344	13.3	2.028	14.2
363	.09676	.07952	.02688	12.76	1.535	13.75
364	.09686	.08468	.05862	16.5	2.344	18.13
365	.07937	.05696	.02181	13.4	1.036	14.73
366	.0915	.1131	.09799	20.44	4.218	24.31
367	.09905	.1669	.1641	20.2	7.128	24.19
368	.09231	.07175	.04392	12.21	1.874	14.29
369	.09384	.08562	.1168	21.71	7.733	30.75
370	.1063	.1954	.2448	22.01	7.561	27.66
371	.09742	.1497	.1811	16.35	2.972	19.38
372	.07963	.06934	.03393	15.19	1.338	16.2
373	.1001	.1515	.1932	21.37	2.407	22.69
374	.09446	.1076	.1527	20.64	4.119	25.37
375	.08302	.06374	.02556	13.69	1.372	14.84
376	.0988	.1438	.06651	16.17	1.349	16.97
377	.09073	.166	.228	10.57	2.363	10.85

378	.07517	.04726	.01271	13.46	1.4	14.69
379	.08268	.07548	.04249	13.66	1.101	14.54
380	.1216	.2154	.1689	11.08	1.719	13.24
381	.1237	.1111	.079	11.27	1.809	12.84
382	.07987	.07079	.03546	11.04	1.281	12.09
383	.06935	.1073	.07943	12.05	1.778	12.57
384	.1042	.1297	.05892	12.39	2.117	14.18
385	.08363	.08575	.05077	13.28	1.592	14.24
386	.08682	.06636	.0839	14.6	2.914	15.79
387	.08108	.07823	.06839	12.21	2.097	13.13
388	.07026	.04831	.02045	13.88	1.709	15.51
389	.08365	.1114	.1007	11.27	2.569	12.04
390	.101	.1318	.1856	19.55	5.383	20.82
391	.09996	.07542	.01923	10.26	1.348	11.38
392	.1039	.07428	0	8.734	3.167	10.17
393	.116	.1562	.1891	15.49	4.675	21.2
394	.1167	.2087	.281	21.61	4.158	26.23
395	.1029	.09758	.04783	12.1	1.869	13.56
396	.08045	.05361	.02681	14.06	1.237	14.92
397	.1059	.1147	.0858	13.51	1.513	14.8
398	.08044	.08895	.0739	12.8	2.668	13.74
399	.07741	.04768	.02712	11.06	1.263	12.68
400	.09087	.06232	.02853	11.8	2.225	13.45
401	.123	.2576	.3189	17.91	3.123	20.8
402	.08872	.05242	.02606	11.93	1.649	13.8
403	.07351	.07899	.04057	12.96	2.397	14.13
404	.09879	.08836	.03296	12.94	.9975	13.86
405	.08682	.04571	.02109	12.34	2.602	13.18
406	.1004	.0746	.04944	10.94	3.018	12.4
407	.09495	.08501	.055	16.14	1.729	17.71
408	.07551	.08316	.06126	12.85	2.552	14.4
409	.1036	.1304	.1201	17.99	3.061	21.08
410	.08685	.06526	.03211	12.27	2.079	14.1
411	.08858	.05313	.02783	11.36	1.359	13.05
412	.1077	.07804	.03046	11.04	1.342	12.41
413	.07969	.06053	.03735	9.397	1.174	9.965
414	.08515	.1025	.06859	14.99	2.31	16.76
415	.0832	.04605	.04686	15.13	3.043	17.26
416	.09773	.0812	.02555	11.89	1.93	13.05
417	.1044	.06159	.02047	9.405	2.759	10.85
418	.112	.1571	.1522	15.5	9.424	23.17
419	.08785	.05794	.0236	12.7	1.527	13.65

420	.1018	.05978	.008955	11.16	1.968	12.36
421	.08546	.07722	.05485	11.57	2.206	13.07
422	.1031	.1836	.145	14.69	4.795	16.46
423	.1088	.1168	.07097	11.61	1.667	12.64
424	.09057	.1147	.09657	13.66	1.804	15.14
425	.1075	.08333	.008934	9.742	4.607	11.21
426	.08117	.03912	.00247	10.03	1.184	11.11
427	.09816	.1013	.06335	10.48	2.564	12.13
428	.08801	.05743	.03614	10.8	2.24	12.76
429	.08151	.03834	.01369	11.13	.968	11.68
430	.07896	.04522	.01402	12.72	2.109	13.82
431	.09947	.2225	.2733	14.9	3.466	16.35
432	.1054	.1316	.07741	12.4	2.204	12.88
433	.1133	.1489	.2133	20.18	3.008	22.03
434	.1018	.1389	.1594	18.82	4.493	22.66
435	.08924	.07074	.03346	14.86	1.612	16.31
436	.106	.1133	.1126	13.98	1.602	17.04
437	.09136	.07883	.01797	12.87	2.597	14.45
438	.08458	.05895	.03534	14.04	2.644	15.66
439	.08684	.0633	.01342	13.85	2.331	15.63
440	.07966	.05581	.02087	14.02	1.606	14.91
441	.08915	.1113	.09457	10.97	2.806	12.36
442	.08331	.1109	.1204	17.27	3.283	20.38
443	.08817	.06718	.01055	13.78	2.235	15.27
444	.08142	.04462	.01993	10.57	1.277	10.94
445	.08947	.1232	.109	18.03	1.921	20.38
446	.103	.09218	.05441	11.99	1.865	12.98
447	.09997	.1314	.1698	17.75	2.873	21.53
448	.09179	.0889	.04069	14.8	1.482	16.43
449	.08388	.078	.08817	14.53	1.994	16.3
450	.09684	.1175	.1572	21.1	4.542	25.68
451	.06613	.1064	.08777	11.87	1.955	12.79
452	.1032	.09871	.1655	19.59	2.916	21.44
453	.08437	.0645	.04055	12	1.516	13.09
454	.1099	.09242	.06895	14.53	2.143	15.8
455	.08583	.0543	.02966	12.62	1.116	14.34
456	.09245	.07426	.02819	13.38	2.287	15.05
457	.09357	.08574	.0716	11.63	2.15	13.12
458	.08791	.05205	.02772	13.21	1.314	14.35
459	.08369	.05073	.01206	13	1.657	14.34
460	.07984	.04626	.01541	9.755	1.243	10.67
461	.09898	.111	.1007	17.08	6.051	22.96

462	.1084	.1988	.3635	27.42	18.65	36.04
463	.06995	.05223	.03476	14.4	1.727	15.4
464	.08508	.05855	.03367	11.6	1.303	12.77
465	.07466	.05994	.04859	13.17	1.236	14.9
466	.08284	.1223	.101	13.24	3.369	15.44
467	.08675	.1089	.1085	13.14	2.312	14.8
468	.08311	.05428	.01479	9.668	2.275	11.15
469	.09289	.2004	.2136	17.6	5.801	21.57
470	.1175	.1483	.102	11.62	3.027	13.36
471	.08946	.06258	.02948	9.667	2.569	11.14
472	.08752	.06	.02367	12.04	4.099	13.6
473	.08098	.08549	.05539	14.92	1.826	17.18
474	.07699	.03398	0	12.27	2.884	13.45
475	.1007	.1069	.05115	10.88	1.301	11.94
476	.0904	.08269	.05835	12.83	1.195	14.09
477	.08931	.1108	.05063	14.2	2.749	16.45
478	.06828	.05319	.02224	13.9	1.392	15.14
479	.1046	.08228	.05308	11.49	1.567	12.4
480	.1026	.1893	.2236	16.25	3.07	17.39
481	.09087	.07838	.02916	12.16	1.678	13.34
482	.07991	.05326	.02995	13.9	2.056	16.41
483	.1071	.1155	.05786	13.47	1.102	14.83
484	.0995	.07957	.04548	13.7	1.564	14.96
485	.1043	.1299	.1191	15.73	1.143	17.01
486	.09514	.1511	.1544	12.45	5.004	13.78
487	.08641	.06698	.05192	14.64	1.471	16.46
488	.1089	.1448	.2256	19.44	3.631	23.96
489	.1128	.09263	.04279	11.68	2.554	13.32
490	.07497	.07112	.03649	16.69	1.775	19.18
491	.08192	.052	.01714	12.25	1.577	14.17
492	.07838	.06217	.04445	17.85	3.163	19.82
493	.1001	.1289	.117	18.01	5.353	21.53
494	.07372	.04043	.007173	12.46	2.108	13.19
495	.07335	.05275	.018	13.16	2.326	14.5
496	.09587	.08345	.06824	14.87	1.596	16.01
497	.1076	.1334	.08017	12.65	1.696	14.38
498	.08928	.0763	.03609	12.47	1.253	14.06
499	.1012	.1317	.1491	18.49	4.851	22.75
500	.1085	.1644	.2188	20.59	4.206	23.86
501	.09883	.1364	.07721	15.04	2.304	16.76
502	.1162	.1681	.1357	13.82	2.974	16.01
503	.1158	.1085	.05928	12.54	1.566	13.57

504	.09342	.1275	.1676	23.09	9.635	30.79
505	.1634	.2239	.0973	9.268	3.014	10.28
506	.1255	.2204	.1188	9.676	1.787	10.6
507	.1096	.1152	.08175	12.22	.9857	13.16
508	.1194	.1071	.04063	11.06	1.318	11.69
509	.09427	.06712	.05526	16.3	1.146	17.32
510	.1183	.187	.203	15.46	2.937	17.11
511	.08099	.09661	.06726	11.74	1.345	12.45
512	.08472	.05016	.03416	14.81	1.677	15.61
513	.1106	.1469	.1445	13.4	3.093	16.41
514	.09832	.08918	.08222	14.58	2.561	16.76
515	.09215	.08597	.07486	15.05	2.63	17.58
516	.1049	.08499	.04302	11.34	1.491	12.47
517	.1068	.1248	.1569	18.31	3.218	21.86
518	.1037	.131	.1411	19.89	3.654	23.73
519	.1218	.1661	.04825	12.88	3.176	15.05
520	.1125	.1117	.0388	12.75	2.495	14.45
521	.1371	.1225	.03332	9.295	2.388	10.57
522	.103	.2106	.231	24.63	7.05	29.92
523	.08511	.04413	.005067	11.26	.9812	11.93
524	.09916	.107	.05385	13.71	2.284	15.11
525	.09492	.08419	.0233	9.847	1.976	11.24
526	.1036	.07632	.02565	8.571	1.069	9.473
527	.1075	.1138	.04201	13.46	1.443	15.35
528	.09003	.06307	.02958	12.34	.7714	13.61
529	.1248	.09755	.101	13.94	4.091	14.62
530	.11	.09009	.03781	12.07	1.714	13.45
531	.1073	.09713	.05282	11.75	3.149	13.5
532	.1016	.09453	.042	11.67	1.393	13.35
533	.09277	.07255	.01752	13.68	1.373	15.85
534	.09156	.1313	.1523	20.47	5.168	23.23
535	.09687	.09752	.05263	10.96	1.165	11.62
536	.1046	.1739	.2085	20.55	4.706	24.3
537	.1038	.1154	.1463	14.27	1.895	15.29
538	.1236	.1552	.04515	11.69	2.158	12.98
539	.08098	.04878	0	7.729	2.492	9.077
540	.08668	.1199	.09252	7.691	1.445	8.678
541	.09984	.112	.06737	11.54	1.628	12.26
542	.08837	.123	.1009	14.47	2.615	16.22
543	.08275	.07214	.04105	14.74	2.177	16.51
544	.08671	.06877	.02987	13.21	1.539	14.37
545	.09578	.1018	.03688	13.87	2.076	15.05

546	.09246	.06747	.02974	13.62	2.066	15.35
547	.09434	.04994	.01012	10.32	1.356	11.25
548	.08877	.08066	.04358	10.26	.9887	10.83
549	.08491	.0503	.02337	9.683	2.054	10.93
550	.08192	.06602	.01548	10.82	3.564	13.03
551	.07431	.04227	0	10.86	2.115	11.66
552	.09566	.08194	.04824	11.13	1.994	12.02
553	.08276	.04234	.01997	12.77	1.477	13.87
554	.0924	.05605	.03996	9.333	2.121	9.845
555	.08123	.05824	.06195	12.88	1.502	13.89
556	.0903	.07658	.05999	10.29	1.437	10.84
557	.1003	.07504	.005025	10.16	1.648	10.65
558	.08123	.04971	0	9.423	3.618	10.49
559	.08473	.133	.1029	14.59	2.224	15.48
560	.09261	.1021	.1112	11.51	1.936	12.48
561	.09929	.1126	.04462	14.05	2.888	15.3
562	.07449	.03558	0	11.2	2.041	11.92
563	.1048	.2087	.255	15.22	2.362	17.52
564	.1099	.2236	.3174	20.92	8.758	24.29
565	.111	.1159	.2439	21.56	7.673	25.45
566	.0978	.1034	.144	20.13	5.203	23.69
567	.08455	.1023	.09251	16.6	3.425	18.98
568	.1178	.277	.3514	20.6	5.772	25.74
569	.05263	.04362	0	7.76	2.548	9.456

Περίληψη

Σκοπός της παρούσας πτυχιακής είναι η μελέτη των διαταραχών των ιδιοτιμών ενός συμμετρικού πίνακα. Συγκεκριμένα, οι συμμετρικοί πίνακες που μας απασχόλησαν είναι οι πίνακες συνδιακύμανσης οι οποίοι είναι εξ' ορισμού θετικά ημιορισμένοι. Στην πολυμεταβλητή στατιστική η εκτίμηση των πινάκων συνδιακύμανσης και συσχέτισης είναι απαραίτητη.

Οι μέθοδοι που αναπτύχθηκαν για τον υπολογισμό πινάκων συνδιακύμανσης που έχουν προέλθει από ελλιπή δεδομένα είναι: μέθοδος παρατηρήσεων με ελλιπή δεδομένα, μέθοδος Expectation-maximization algorithm (EM) και μέθοδος διαγραφής ανά ζεύγη.

Χρησιμοποιώντας τη μέθοδο διαγραφής ανά ζεύγη, κατά την ανάλυση, ίσως ο πίνακας συνδιακύμανσης που θα προκύψει να μην είναι θετικά ημιορισμένος. Οι αόριστοι αυτοί πίνακες διαταράσσονται με τη χρήση των ακόλουθων μεθόδων: φασματικής ανάλυσης, μεταφοράς και μέσω πίνακα συσχέτισης.

Η σύγκριση των μεθόδων διαταραχής γίνεται με τη χρήση νόρμας Frobenius και της γραμμικής παλινδρόμησης.

Τα αποτελέσματα της σύγκρισης των μεθόδων διαταραχής έδειξαν ότι το μέγεθος του συνόλου δεδομένων και το pvalue των μεταβλητών επηρεάζει την επιλογή της μεθόδου.

Για τα παραπάνω υλοποιήθηκαν κώδικες σε Stata.

Λέξεις κλειδιά: θετικά ημιορισμένος πίνακας, πίνακας συνδιακύμανσης-συσχέτισης, νόρμες, φασματική ανάλυση, μέθοδος διαγραφής ανά ζεύγη, γραμμική παλινδρόμηση

Abstract

The aim of this thesis is to study the perturbation of an indefinite matrix into a positive semidefinite matrix. More specifically, we are interested in the symmetrical covariance matrices which are almost always positive semidefinite. In multivariate statistics, the estimation of covariance and correlation matrices is necessary.

The methods which are implemented to estimate covariance matrix from a full dataset are: listwise method, expectation-maximization algorithm (EM), pairwise method.

Using the pairwise method in order to analyse, the variance-covariance matrix which is derived, may not be a semidefinite matrix. These indefinite matrices are perturbed using the following methods: spectral decomposition, transformation and through correlation matrix.

The perturbation methods' comparison has been conducted using the Frobenius norm and the linear regression.

The comparison's results of the perturbation methods conclude to the dependence of the method's choice on the dataset size and the pvalue of the variables.

For the above stata code are structured.

Keywords: semidefinite matrix, covariance-correlation matrix, norm, spectral decomposition, pairwise method, linear regression