ΚΕΦΑΛΑΙΟ 3

ΧΩΡΙΚΗ ΑΥΤΟΣΥΣΧΕΤΙΣΗ

3.1 Εισαγωγή

Η χωρική αυτοσυσχέτιση (*spatial autocorrelation*) αποτελεί την ασθενέστερή μορφή της χωρικής εξάρτησης και είναι το χαρακτηριστικό της που εμφανίζεται στην πράξη και που μπορεί να μελετηθεί από τον ερευνητή. Χωρική αυτοσυσχέτιση σημαίνει ότι υπάρχει τάση στις παρατηρήσεις μιας μεταβλητής να συγκεντρώνονται στο γεωγραφικό χώρο δημιουργώντας συστάδες (*clusters*) τιμών και να μην κατανέμονται με τυχαίο τρόπο. Δηλαδή, πρόκειται για συσχέτιση μιας μεταβλητής με τιμές της ίδιας μεταβλητής σε γειτονικές περιοχές κατά κάποιο τρόπο όπως συμβαίνει και στην ανάλυση χρονοσειρών με τη χρονική αυτοσυσχέτιση όπου οι τιμές της χρονοσειράς συσχετίζονται με τις τιμές της σε προηγούμενες χρονικές περιόδους. Ωστόσο, ενώ στις χρονοσειρές η συσχέτιση είναι μόνο προς μία κατεύθυνση στη χωρική αυτοσυσχέτιση η συσχέτιση εμφανίζεται προς πολλές κατευθύνσεις. Ο πρώτος νόμος της γεωγραφίας του Tobler (1970) σύμφωνα με τον οποίον "καθετί στο χώρο συσχετίζεται με κάτι άλλο αλλά τα κοντινότερα πράγματα έχουν μεγαλύτερη συσχέτιση από αυτά που είναι πιο απομακρυσμένα" εκφράζει το φαινόμενο της χωρικής αυτοσυσχέτισης.

Η εμφάνιση χωρικής αυτοσυσχέτισης στις τιμές μιας μεταβλητής παραβιάζει την υπόθεση του τυχαίου δείγματος και προκαλεί προβλήματα σε κάθε στατιστική ανάλυση. Η κλασική στατιστική και οικονομετρική μεθοδολογία όχι μόνο δεν αντιμετωπίζει τα προβλήματα αυτά αλλά πολλές φορές δεν τα εντοπίζει. Η παρουσία χωρικής αυτοσυσχέτισης καθιστά αναξιόπιστο ακόμη και το διάστημα εμπιστοσύνης για τη μέση τιμή της μεταβλητής καθώς εάν δεν ληφθεί υπ' όψιν η διακύμανση του εκτιμητή υποεκτιμάται. Ανάλογα προβλήματα προκύπτουν και κατά την εκτίμηση οικονομετρικών υποδειγμάτων διότι τα σφάλματά τους δεν θα είναι ασυσχέτιστα. Στο κεφάλαιο αυτό παρουσιάζεται η μεθοδολογία που χρησιμοποιείται τις περισσότερες φορές για την εκτίμηση και τον έλεγχο της ύπαρξης χωρικής αυτοσυσχέτισης στις τιμές μιας μεταβλητής.

3.2 Διακρίσεις της χωρικής αυτοσυσχέτισης

Η χωρική αυτοσυσχέτιση εμφανίζεται στην πράξη ως συστάδες στις τιμές της εξεταζόμενης μεταβλητής όταν αυτές απεικονίζονται επάνω στο χάρτη της υπό μελέτης γεωγραφικής περιοχής. Όπως ισχύει και στο συντελεστή γραμμικής συσχέτισης του Pearson υπάρχουν οι εξής τρείς περιπτώσεις:

Α) Θετική χωρική αυτοσυσχέτιση

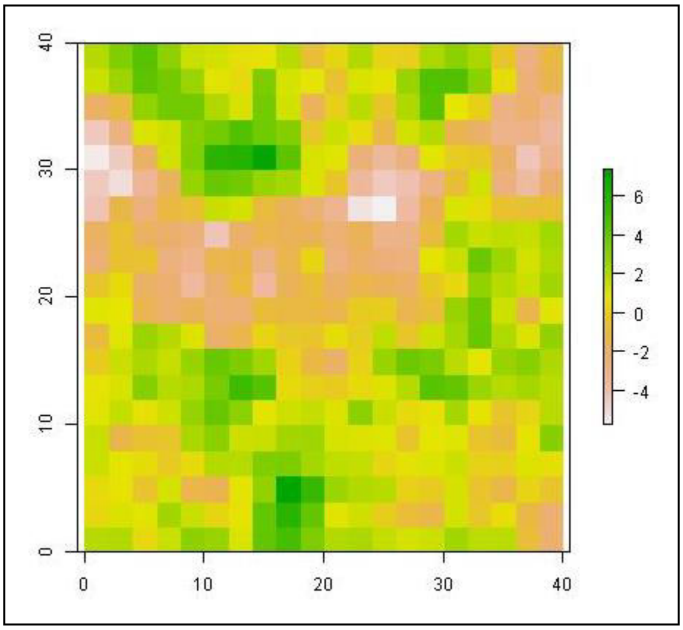
Όταν υπάρχει θετική χωρική αυτοσυσχέτιση οι υψηλές τιμές της μεταβλητής βρίσκονται κοντά σε υψηλές τιμές και οι χαμηλές τιμές κοντά σε χαμηλές τιμές σχηματίζοντας στο χάρτη συστάδες υψηλών και χαμηλών τιμών.

Β) Αρνητική χωρική αυτοσυσχέτιση

Στην περίπτωση της αρνητικής χωρικής αυτοσυσχέτισης υψηλές τιμές της μεταβλητής βρίσκονται κοντά σε χαμηλές τιμές και οι χαμηλές τιμές κοντά σε υψηλές τιμές εμφανίζοντας στο χάρτη ένα πρότυπο που θυμίζει σκακιέρα.

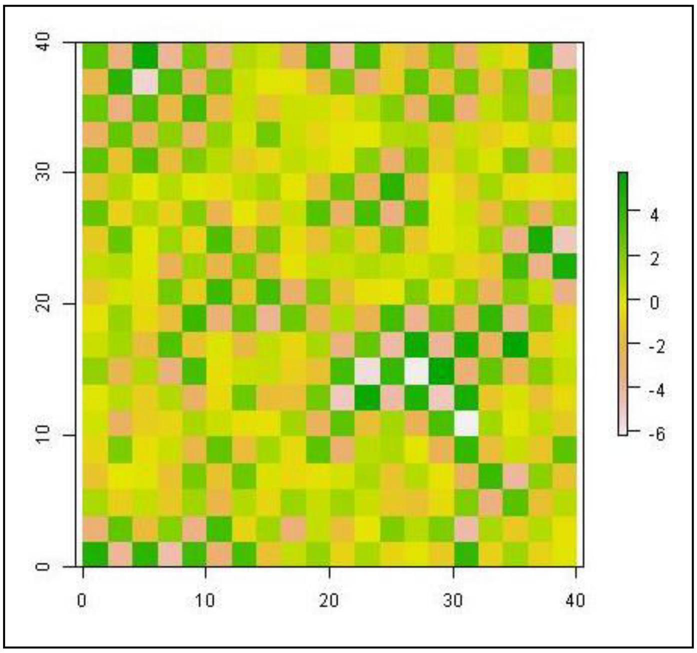
Γ) Απουσία χωρικής αυτοσυσχέτισης

Όταν δεν υπάρχει χωρική αυτοσυσχέτιση στις τιμές της μεταβλητής αυτές θα κατανέμονται στο χώρο με τυχαίο τρόπο και η χαρτογράφησή τους δεν θα εμφανίζει κάποιο ιδιαίτερο πρότυπο.

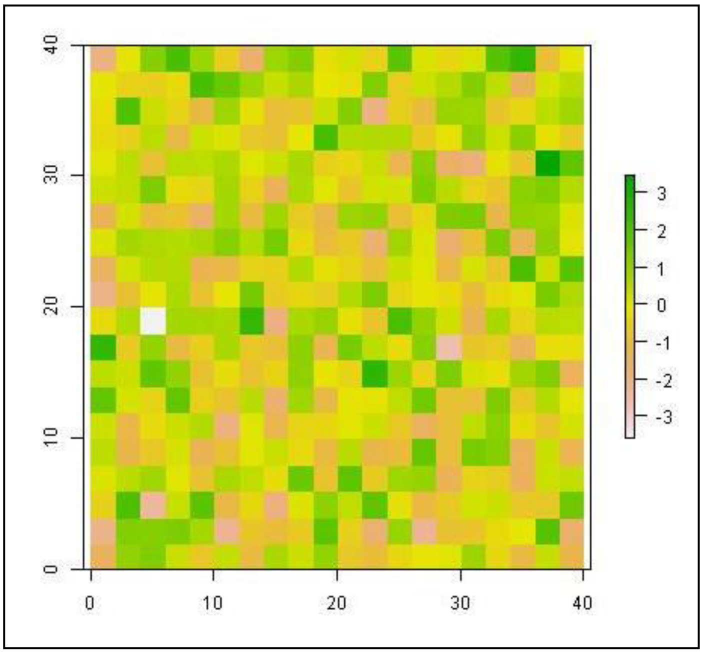


Διάγραμμα 3.1
Θετική χωρική αυτοσυσχέτιση

Στο Διάγραμμα 3.1 παρουσιάζεται ένα ομαλό πλέγμα 400 υποθετικών χωρικών μονάδων επάνω στο οποίο έχουν χαρτογραφηθεί οι τιμές μιας μεταβλητής με ισχυρή θετική χωρική αυτοσυσχέτιση. Εύκολα μπορούν να παρατηρηθούν οι σχηματισμοί συστάδων υψηλών τιμών και συστάδων χαμηλών τιμών όπως φαίνεται με τη βοήθεια της χρωματικής διαβάθμισης του υπομνήματος.

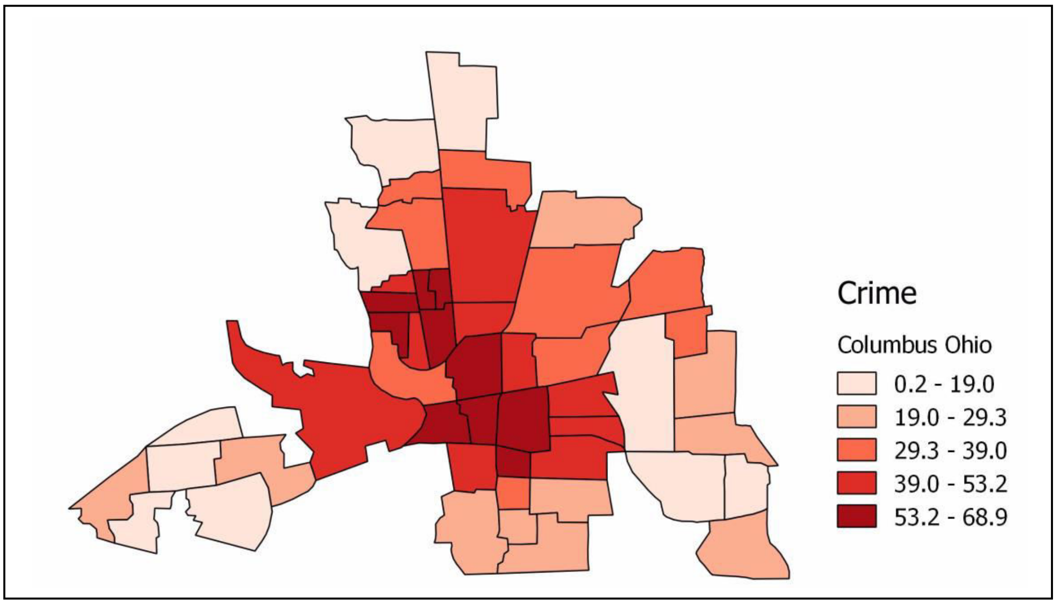


Διάγραμμα 3.2
Αρνητική χωρική αυτοσυσχέτιση



Διάγραμμα 3.3
Απουσία χωρικής αυτοσυσχέτισης

Στο Διάγραμμα 3.2 παρουσιάζεται ένα ομαλό πλέγμα επάνω στο οποίο έχουν χαρτογραφηθεί οι τιμές μιας μεταβλητής με ισχυρή αρνητική χωρική αυτοσυσχέτιση. Όπως διαπιστώνεται, χαμηλές τιμές της μεταβλητής γειτονεύουν με υψηλές και αντίστροφα προκαλώντας την εμφάνιση του προτύπου της σκακιέρας. Στο Διάγραμμα 3.3 η μεταβλητή δεν έχει χωρική αυτοσυσχέτιση ούτε θετική ούτε αρνητική. Οι τιμές της απεικονίζονται με τυχαίο τρόπο και δεν παρατηρείται κάποιο χωρικό πρότυπο.



Διάγραμμα 3.4
Θετική χωρική αυτοσυσχέτιση στα συμβάντα εγκληματικότητας στην πρωτεύουσα του Οχάιο Columbus στις ΗΠΑ

Ο χάρτης στο Διάγραμμα 3.4 απεικονίζει τις καταγραφές διαρρήξεων οικιών και κλοπών οχημάτων ανά χιλίους κατοίκους που έχουν καταγραφεί το έτος 1980 στην πρωτεύουσα του Οχάιο Columbus των ΗΠΑ.[1] Η βασική διαπίστωση από το χάρτη είναι η ύπαρξη θετικής χωρικής αυτοσυχέτισης στην εγκληματικότητα. Τα περισσότερα περιστατικά βρίσκονται μαζί συγκεντρωμένα σχηματίζοντας συστάδα υψηλών τιμών στο κέντρο της πόλης. Αντίθετα, τα λιγότερα περιστατικά καταγράφονται περιμετρικά της πόλης σχηματίζοντας συστάδες χαμηλών τιμών.

[1] Τα δεδομένα εγκληματικότητας του Columbus έχουν αναφερθεί πολλές φορές τόσο σε εργασίες όσο και σε εγχειρίδια χωρικής ανάλυσης. Παρουσιάστηκαν αρχικά στο βιβλίο του Luc Anselin “Spatial Econometrics: Methods and Models” και συμπεριλαμβάνονται ως παράδειγμα σε πακέτα της γλώσσας R αλλά και ως συνοδευτικά αρχεία στο πρόγραμμα χωρικής ανάλυσης GeoDa.

3.3 Ο συντελεστής χωρικής αυτοσυσχετισης *I* του Moran

Ο πιο παλιός και πιο συχνά χρησιμοποιούμενος τρόπος για τη μέτρηση της ύπαρξης χωρικής αυτοσυσχέτισης σε μια μεταβλητή y είναι με τον υπολογισμό του συντελεστή χωρικής αυτοσυσχέτισης I που έχει προταθεί από τον Moran (1948 και 1950). Ο συντελεστή I ορίζεται ως:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(y_i - \bar{y}\right)\left(y_j - \bar{y}\right)}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

όπου n είναι ο αριθμός των γεωγραφικών περιοχών, $\bar{y}$ η μέση τιμή της μεταβλητής y και w_{ij} το αντίστοιχο στοιχείο της μήτρας χωρικών σταθμίσεων **W**. Όταν η μήτρα χωρικών σταθμίσεων χρησιμοποιείται στην τυποποιημένη κατά γραμμή μορφή της θα ισχύει ότι $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} = n$ και ο συντελεστής I γράφεται ισοδύναμα:

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(y_i - \bar{y}\right)\left(y_j - \bar{y}\right)}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

Ο συντελεστή αυτός έχει ομοιότητες με το συντελεστή γραμμικής συσχέτισης r του Pearson. Ειδικότερα, ο αριθμητής στο προηγούμενο πηλίκο θα μπορούσε να ερμηνευθεί ως η συνδιακύμανση της μεταβλητής y με τη χωρική της υστέρηση ενώ ο παρανομαστής είναι το άθροισμα τετραγώνων των αποκλίσεων των τιμών της μεταβλητής από τη μέση τιμή της.

Εάν υποτεθεί ότι δεν υπάρχει χωρική αυτοσυσχέτιση τότε ο συντελεστή I θα έχει αναμενόμενη τιμή:

$$E(I) = \frac{-1}{n-1}$$

η οποία εξαρτάται αποκλειστικά από το μέγεθος του δείγματος.

Ο συντελεστής I ερμηνεύεται ως εξής:

- Εάν $I > E(I)$, τότε υπάρχει θετική χωρική αυτοσυσχέτιση.[2] Αυτό σημαίνει ότι οι παρατηρήσεις της μεταβλητής y σχηματίζουν συστάδες στο χώρο έτσι ώστε οι μεγάλες τιμές να βρίσκονται κοντά στις μεγάλες τιμές και οι μικρές τιμές κοντά στις μικρές.
- Όταν $I < E(I)$ υπάρχει αρνητική χωρική αυτοσυσχέτιση και στο χώρο θα σχηματίζονται συστάδες μεγάλων και μικρών τιμών και αντίθετα εμφανίζοντας στο χάρτη ένα πρότυπο που μοιάζει με σκακιέρα.
- Όταν $I = E(I)$ δεν υπάρχει χωρική αυτοσυσχέτιση και οι παρατηρήσεις κατανέμονται στο χώρο με τυχαίο τρόπο.

3.3.1 Ο συντελεστής χωρικής αυτοσυσχετισης *I* του Moran με μήτρες

Χρησιμοποιώντας μήτρες ο συντελεστή I του Moran ορίζεται ως εξής:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{(\mathbf{y}-\bar{\mathbf{y}})' \mathbf{W} (\mathbf{y}-\bar{\mathbf{y}})}{(\mathbf{y}-\bar{\mathbf{y}})' (\mathbf{y}-\bar{\mathbf{y}})}$$

όπου $\mathbf{y}$ είναι το $(n \times 1)$ διάνυσμα των τιμών της μεταβλητής y, $\bar{\mathbf{y}}$ ένα $(n \times 1)$ διάνυσμα που σε κάθε γραμμή περιέχει τη μέση τιμή της y και $\mathbf{W}$ η $(n \times n)$ μήτρα χωρικών σταθμίσεων. Όταν η μήτρα χωρικών σταθμίσεων είναι τυποποιημένη ανά γραμμή ο συντελεστή I θα γράφεται ισοδύναμα ως εξής:

$$I = \frac{(\mathbf{y}-\bar{\mathbf{y}})' \mathbf{W} (\mathbf{y}-\bar{\mathbf{y}})}{(\mathbf{y}-\bar{\mathbf{y}})' (\mathbf{y}-\bar{\mathbf{y}})}$$

[2] Πολλές φορές στη βιβλιογραφία θεωρείται ότι υπάρχει θετική χωρική αυτοσυσχέτιση όταν $I > 0$ και αρνητική αυτοσυσχέτιση όταν $I < 0$. Η προσέγγιση αυτή στην πραγματικότητα δεν διαφέρει από την παρουσίαση αυτής της ενότητας καθώς είναι φανερό ότι ουσιαστικά ακόμη και για μικρά n το $E(I)$ θα είναι πολύ κοντά στο μηδέν.

3.3.2 Στατιστική συμπερασματολογία για το συντελεστή *I* του Moran

Η στατιστική συμπερασματολογία για το συντελεστή χωρικής αυτοσυσχέτισης *I* του Moran μπορεί να διεξαχθεί με τρείς διαφορετικές προσεγγίσεις που αφορούν τον υπολογισμό των ροπών του. Οι προσεγγίσεις αυτές είναι η υπόθεση της κανονικότητας, η υπόθεση της τυχαιοποίησης και η εφαρμογή μεταθέσεων.

Α) Κανονικότητα (*Normality*)

Οι Cliff και Ord (1981) υπολόγισαν τις ροπές της *I* υποθέτοντας τη συνθήκη της κανονικότητας. Ειδικότερα, εάν η μεταβλητή *y* ακολουθεί την κανονική κατανομή. τότε η διακύμανση της *I* θα είναι:

$$Var(I) = E(I^2) - E(I)^2$$

όπου

$$E(I) = \frac{-1}{n-1}$$

και

$$E(I^2) = \frac{n^2 S_1 - nS_2 + 3{S_0}^2}{{S_0}^2\left(n-1\right)\left(n+1\right)}$$

με

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \ ,$$

$$S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(w_{ij} + w_{ji}\right)^2 ,$$

και

$$S_2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2$$

Δηλαδή, η διακύμανση θα εξαρτάται μόνο από το μέγεθος του δείγματος και τα στοιχεία της μήτρας χωρικών σταθμίσεων. Κατά συνέπεια, υποθέτοντας κανονική κατανομή, ο έλεγχος της μηδενικής υπόθεσης της απουσίας χωρικής αυτοσυσχέτισης μπορεί να γίνει με τον μετασχηματισμό:

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}}$$

που θα ακολουθεί την τυποποιημένη κανονική κατανομή.

B) Τυχαιοποίηση (*Randomization*)

Όταν η μεταβλητή y δεν ακολουθεί κανονική κατανομή ο έλεγχος της μηδενικής υπόθεσης για την απουσία χωρικής αυτοσυσχέτισης θα μπορούσε να πραγματοποιηθεί κάνοντας την υπόθεση της τυχαιοποίησης (*randomization*), δηλαδή θεωρώντας την εκτιμηθείσα τιμή της I σε σχέση με όλες τις δυνατές τιμές που θα μπορούσε να λάβει εάν υποθέτονταν όλες οι δυνατές μεταθέσεις της y σε όλες τις γεωγραφικές περιοχές. Ο Ord (1980) έδειξε ότι εάν υποτεθεί το σύνολο των $n!$ ισοπίθανων τυχαίων μεταθέσεων της y τότε υπό τη μηδενική υπόθεσης της απουσίας χωρικής αυτοσυσχέτισης και δεδομένου του παρατηρηθέντος δείγματος αλλάζει η δεύτερη ροπή της I η οποία θα είναι:

$$E(I^2) = \frac{n\left[\left(n^2-3n+3\right)S_1 - nS_2 + 3S_0{}^2\right] - b_2\left[\left(n^2-n\right)S_1 - 2nS_2 + 6S_0{}^2\right]}{\left(n-1\right)\left(n-2\right)\left(n-3\right)S_0{}^2}$$

με

$$b_2 = \frac{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^4}{\left[\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2\right]^2} n = \frac{\dfrac{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^4}{n}}{\left[\dfrac{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}{n}\right]^2}$$

Δηλαδή, στην περίπτωση αυτή η διακύμανση θα εξαρτάται και από τις τιμές του δείγματος και από τα στοιχεία της μήτρας χωρικών σταθμίσεων. Στον υπολογισμό της δεύτερης ροπής b_2 είναι ο συντελεστής κύρτωσης που έχει σκοπό τη διόρθωση της διακύμανσης ως προς την κανονικότητα. Συγκεκριμένα, όσο με βάση το δείγμα η κατανομή της στατιστικής I απομακρύνεται από την υπόθεση της κανονικότητας η διακύμανσή της θα αυξάνεται. Κατά συνέπεια, η διακύμανση της I θα υπολογίζεται ως

$$Var(I) = E(I^2) - E(I)^2$$

όπου

$$E(I) = \frac{-1}{n-1}$$

Εφόσον το δείγμα είναι αρκετά μεγάλο η I μετασχηματίζεται και πάλι σε τυποποιημένη κανονική μεταβλητή και θα διεξάγεται ο έλεγχος με τη στατιστική:

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}}$$

Γ) Εφαρμογή μεταθέσεων (*Permutations*)

Σύμφωνα με αυτή την προσέγγιση πραγματοποιούνται τυχαίες μεταθέσεις των τιμών της y στις χωρικές μονάδες και επαναϋπολογίζεται η τιμή της I πολλές φορές για να παραχθεί μια κατανομή αναφοράς και να συγκριθεί με αυτή η τιμή της I που υπολογίστηκε με βάση το διαθέσιμο δείγμα. Η διαδικασία γίνεται με τη βοήθεια λογισμικού και περιλαμβάνεται και στο λογισμικό χωρικής ανάλυσης GeoDa.

3.3.3 Σχόλια για το συντελεστή *I* του Moran

1) Ο συντελεστής I του Moran συνήθως ερμηνεύεται όπως ο συντελεστής γραμμικής συσχέτισης του Pearson.
2) Σε αντίθεση με ότι ισχύει για το συντελεστή γραμμικής συσχέτισης, ο συντελεστής I μπορεί να λάβει τιμές εκτός του εύρους [-1, 1].
3) Για τυποποιημένες ανά γραμμή μήτρες χωρικών σταθμίσεων **W** το εύρος του συντελεστή I θα μπορούσε να θεωρηθεί ότι είναι το διάστημα [-1, 1]. Για μη τυποποιημένες ανά γραμμή μήτρες χωρικών σταθμίσεων δεν υπάρχει καμία εγγύηση ότι το προηγούμενο διάστημα ισχύει.
4) Το εύρος τιμών του I εξαρτάται αποκλειστικά από τη μήτρα χωρικών σταθμίσεων **W** που χρησιμοποιείται στην ανάλυση και όχι από τις τιμές της μεταβλητής, δηλαδή θα διαφέρει σε κάθε γεωγραφική περιοχή μελέτης. Οι de Jong, Sprenger και Veen (1984) έδειξαν ότι οι τιμές της στατιστικής I βρίσκονται στο διάστημα:

$$\frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \lambda_{\min} \leq I \leq \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \lambda_{\max}$$

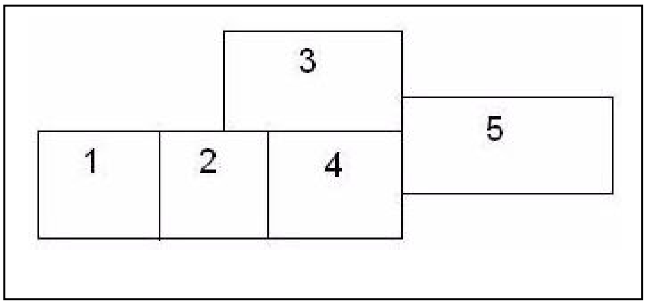
όπου $\lambda_{\min}$ και $\lambda_{\max}$ είναι η ελάχιστη και η μέγιστη ιδιοτιμή της μήτρας:

$$\left(\mathbf{I} - \frac{1}{n}\mathbf{11}'\right)\mathbf{W}\left(\mathbf{I} - \frac{1}{n}\mathbf{11}'\right)$$

στην οποία με $\mathbf{I}$ συμβολίζεται η ($n \times n$) μοναδιαία μήτρα και με $\mathbf{1}$ ένα ($n \times 1$) διάνυσμα από μονάδες. Όλες οι δυνατές τιμές του Ι προκύπτουν από τις ιδιοτιμές της προηγούμενης μήτρας αφού πολλαπλασιαστούν με το $n \Big/ \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$ ή με το 1 για τυποποιημένες μήτρες. Δηλαδή, σε αυτή την περίπτωση οι ιδιοτιμές είναι οι δυνατές τιμές του I.

Παράδειγμα 3.1

Στο Διάγραμμα 3.5 παρουσιάζονται πέντε υποθετικές χωρικές μονάδες και στον Πίνακα 3.1 οι τιμές μιας μεταβλητής y που έχουν καταγραφεί σε κάθε μια από αυτές.



Διάγραμμα 3.5
Πέντε υποθετικές χωρικές μονάδες

Πίνακας 3.1
Τιμές μεταβλητής *y*

Περιοχή i	y_i
1	5
2	6
3	16
4	14
5	14

Η μήτρα χωρικής γειτονίας για τις πέντε χωρικές μονάδες του Διαγράμματος 3.1 με εφαρμογή του κριτηρίου του πύργου είναι:

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Η τυποποιημένη ανά γραμμή μήτρα χωρικών σταθμίσεων είναι:

$$\mathbf{W} = \begin{bmatrix} 0{,}000 & 1{,}000 & 0{,}000 & 0{,}000 & 0{,}000 \\ 0{,}333 & 0{,}000 & 0{,}333 & 0{,}333 & 0{,}000 \\ 0{,}000 & 0{,}333 & 0{,}000 & 0{,}333 & 0{,}333 \\ 0{,}000 & 0{,}333 & 0{,}333 & 0{,}000 & 0{,}333 \\ 0{,}000 & 0{,}000 & 0{,}500 & 0{,}500 & 0{,}000 \end{bmatrix}$$

Το (5×1) διάνυσμα των τιμών της μεταβλητής y είναι:

$$\mathbf{y} = \begin{bmatrix} 5 \\ 6 \\ 16 \\ 14 \\ 14 \end{bmatrix}$$

Ο συντελεστή χωρικής αυτοσυσχέτισης I του Moran θα προκύψει από τη σχέση:

$$I = \frac{(\mathbf{y} - \overline{\mathbf{y}})' \, \mathbf{W} (\mathbf{y} - \overline{\mathbf{y}})}{(\mathbf{y} - \overline{\mathbf{y}})' (\mathbf{y} - \overline{\mathbf{y}})}$$

Αρχικά υπολογίζεται η μέση τιμή της y που είναι:

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{5} y_i}{5} = \frac{5 + 6 + 16 + 14 + 14}{5} = \frac{55}{5} = 11$$

Το (5×1) διάνυσμα $\overline{\mathbf{y}}$ που σε κάθε γραμμή περιέχει τη μέση τιμή της y είναι:

$$\overline{\mathbf{y}} = \begin{bmatrix} 11 \\ 11 \\ 11 \\ 11 \\ 11 \end{bmatrix}$$

Το (5×1) διάνυσμα $\mathbf{y} - \overline{\mathbf{y}}$ με τις αποκλίσεις της y από τη μέση τιμή της είναι:

$$\mathbf{y} - \overline{\mathbf{y}} = \begin{bmatrix} 5 \\ 6 \\ 16 \\ 14 \\ 14 \end{bmatrix} - \begin{bmatrix} 11 \\ 11 \\ 11 \\ 11 \\ 11 \end{bmatrix} = \begin{bmatrix} -6 \\ -5 \\ 5 \\ 3 \\ 3 \end{bmatrix}$$

Το ανάστροφο διάνυσμα $(\mathbf{y} - \overline{\mathbf{y}})'$ θα είναι:

$$(\mathbf{y} - \overline{\mathbf{y}})' = \begin{bmatrix} -6 & -5 & 5 & 3 & 3 \end{bmatrix}$$

Ο αριθμητής στο πηλίκο του συντελεστή I είναι:

$$(\mathbf{y}-\bar{\mathbf{y}})'\mathbf{W}(\mathbf{y}-\bar{\mathbf{y}}) = \begin{bmatrix}-6 & -5 & 5 & 3 & 3\end{bmatrix}\begin{bmatrix}0{,}000 & 1{,}000 & 0{,}000 & 0{,}000 & 0{,}000\\ 0{,}333 & 0{,}000 & 0{,}333 & 0{,}333 & 0{,}000\\ 0{,}000 & 0{,}333 & 0{,}000 & 0{,}333 & 0{,}333\\ 0{,}000 & 0{,}333 & 0{,}333 & 0{,}000 & 0{,}333\\ 0{,}000 & 0{,}000 & 0{,}500 & 0{,}500 & 0{,}000\end{bmatrix}\begin{bmatrix}-6\\-5\\5\\3\\3\end{bmatrix}$$

$$= \begin{bmatrix}\text{-1,667} & \text{-3,333} & 0{,}833 & 1{,}5 & 2{,}667\end{bmatrix}\begin{bmatrix}-6\\-5\\5\\3\\3\end{bmatrix} = 43{,}3333$$

Ο παρανομαστής στο πηλίκο του συντελεστή I είναι:

$$(\mathbf{y}-\bar{\mathbf{y}})'(\mathbf{y}-\bar{\mathbf{y}}) = \begin{bmatrix}-6 & -5 & 5 & 3 & 3\end{bmatrix}\begin{bmatrix}-6\\-5\\5\\3\\3\end{bmatrix} = 104$$

Επομένως ο συντελεστή χωρικής αυτοσυσχέτισης I του Moran θα είναι:

$$I = \frac{(\mathbf{y}-\bar{\mathbf{y}})'\mathbf{W}(\mathbf{y}-\bar{\mathbf{y}})}{(\mathbf{y}-\bar{\mathbf{y}})'(\mathbf{y}-\bar{\mathbf{y}})} = \frac{43{,}3333}{104} = 0{,}41667$$

Η αναμενόμενη τιμή του I είναι:

$$E(I) = \frac{-1}{n-1} = \frac{-1}{5-1} = \frac{-1}{4} = -0{,}25$$

Επειδή $I = 0{,}41667 > E(I) = -0{,}25$ υπάρχει θετική χωρική αυτοσυσχέτιση.

Στη συνέχεια θα παρουσιαστεί η διαδικασία υπολογισμού της διακύμανσης της I και θα πραγματοποιηθεί ο στατιστικός έλεγχος της μηδενικής υπόθεσης της απουσίας χωρικής αυτοσυσχέτισης υποθέτοντας κανονικότητα και τυχαιοποίηση. Εδώ πρέπει να επισημανθεί ότι το δείγμα είναι πολύ μικρό για να ισχύει η κανονική κατανομή άλλα η διαδικασία παρουσιάζεται για λόγους παραδείγματος. Η διακύμανση θα υπολογιστεί από τη σχέση:

$$Var(I) = E(I^2) - E(I)^2$$

Α) Η δεύτερη ροπή της I υποθέτοντας κανονικότητα θα είναι:

$$E(I^2)=\frac{n^2S_1-nS_2+3S_0{}^2}{S_0{}^2(n-1)(n+1)}$$

Οι ποσότητες S_0, S_1 και S_2 υπολογίζονται ως εξής:

$$S_0=\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}=\sum_{i=1}^{5}\sum_{j=1}^{5}w_{ij}=1+1+1+1+1=5=n$$

$$\begin{aligned}
S_1&=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(w_{ij}+w_{ji}\right)^2=\frac{1}{2}\sum_{i=1}^{5}\sum_{j=1}^{5}\left(w_{ij}+w_{ji}\right)^2\\
&=\frac{1}{2}\sum_{i=1}^{5}\left[\left(w_{i1}+w_{1i}\right)^2+\left(w_{i2}+w_{2i}\right)^2+\left(w_{i3}+w_{3i}\right)^2+\left(w_{i4}+w_{4i}\right)^2+\left(w_{i5}+w_{5i}\right)^2\right]\\
&=\frac{1}{2}\{\ \left[\left(w_{11}+w_{11}\right)^2+\left(w_{12}+w_{21}\right)^2+\left(w_{13}+w_{31}\right)^2+\left(w_{14}+w_{41}\right)^2+\left(w_{15}+w_{51}\right)^2\right]\\
&\quad+\left[\left(w_{21}+w_{12}\right)^2+\left(w_{22}+w_{22}\right)^2+\left(w_{23}+w_{32}\right)^2+\left(w_{24}+w_{42}\right)^2+\left(w_{25}+w_{52}\right)^2\right]\\
&\quad+\left[\left(w_{31}+w_{13}\right)^2+\left(w_{32}+w_{23}\right)^2+\left(w_{33}+w_{33}\right)^2+\left(w_{34}+w_{43}\right)^2+\left(w_{35}+w_{53}\right)^2\right]\\
&\quad+\left[\left(w_{41}+w_{14}\right)^2+\left(w_{42}+w_{24}\right)^2+\left(w_{43}+w_{34}\right)^2+\left(w_{44}+w_{44}\right)^2+\left(w_{45}+w_{54}\right)^2\right]\\
&\quad+\left[\left(w_{51}+w_{15}\right)^2+\left(w_{52}+w_{25}\right)^2+\left(w_{53}+w_{35}\right)^2+\left(w_{54}+w_{45}\right)^2+\left(w_{55}+w_{55}\right)^2\right]\ \}
\end{aligned}$$

$$\begin{aligned}
&=\frac{1}{2}\{\ \left[(0+0)^2+(1+0.333)^2+(0+0)^2+(0+0)^2+(0+0)^2\right]\\
&+\left[(0.333+1)^2+(0+0)^2+(0,333+0,333)^2+(0,333+0,333)^2+(0+0)^2\right]\\
&+\left[(0+0)^2+(0,333+0,333)^2+(0+0)^2+(0,333+0,333)^2+(0,333+0,5)^2\right]\\
&+\left[(0+0)^2+(0,333+0,333)^2+(0,333+0,333)^2+(0+0)^2+(0,333+0,5)^2\right]\\
&+\left[(0+0)^2+(0+0)^2+(0,5+0,333)^2+(0,5+0,333)^2+(0+0)^2\right]\ \}
\end{aligned}$$

$$\begin{aligned}
&=\frac{1}{2}\{\ \left[(1.333)^2\right]\\
&+\left[(1.333)^2+(0,666)^2+(0,666)^2\right]\\
&+\left[(0,666)^2+(0,666)^2+(0,833)^2\right]\\
&+\left[(0,666)^2+(0,666)^2+(0,833)^2\right]\\
&+\left[(0,833)^2+(0,833)^2\right]\ \}
\end{aligned}$$

Και τελικά:

$$S_1 = \frac{1}{2}\left[1{,}776889 + 2{,}664001 + 1{,}581001 + 1{,}581001 + 1{,}387778\right] = \frac{1}{2}8{,}99067 = 4{,}5$$

Για την ποσότητα S_2 ισχύει:

$$\begin{aligned}
S_2 &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2 = \sum_{i=1}^{5}\left(\sum_{j=1}^{5} w_{ij} + \sum_{j=1}^{5} w_{ji}\right)^2 \\
&= \left(\sum_{j=1}^{5} w_{1j} + \sum_{j=1}^{5} w_{j1}\right)^2 + \left(\sum_{j=1}^{5} w_{2j} + \sum_{j=1}^{5} w_{j2}\right)^2 + \left(\sum_{j=1}^{5} w_{3j} + \sum_{j=1}^{5} w_{j3}\right)^2 \\
&\quad + \left(\sum_{j=1}^{5} w_{4j} + \sum_{j=1}^{5} w_{j4}\right)^2 + \left(\sum_{j=1}^{5} w_{5j} + \sum_{j=1}^{5} w_{j5}\right)^2
\end{aligned}$$

Μπορεί εύκολα να διαπιστωθεί ότι εντός κάθε παρένθεσης περιέχεται το άθροισμα του συνόλου κάθε γραμμής με το σύνολο της αντίστοιχης στήλης οπότε προκύπτει:

$$\begin{aligned}
S_2 &= \left(1 + 0{,}333\right)^2 + \left(1 + 1 + 0{,}333 + 0{,}333\right)^2 + \left(1 + 0{,}333 + 0{,}333 + 0{,}5\right)^2 \\
&\quad + \left(1 + 0{,}333 + 0{,}333 + 0{,}5\right)^2 + \left(1 + 0{,}333 + 0{,}333\right)^2 \\
&= \left(1{,}333\right)^2 + \left(2{,}666\right)^2 + \left(2{,}166\right)^2 + \left(2{,}166\right)^2 + \left(1{,}666\right)^2 \\
&= 1{,}776889 + 7{,}107556 + 4{,}691556 + 4{,}691556 + 2{,}775556 \\
&= 21{,}05556
\end{aligned}$$

Επομένως, η δεύτερη ροπή του συντελεστή I θα είναι:

$$\begin{aligned}
E(I^2) &= \frac{n^2 S_1 - n S_2 + 3 S_0^{\,2}}{S_0^{\,2}\left(n-1\right)\left(n+1\right)} = \frac{5^2 \cdot 4{,}5 - 5 \cdot 21{,}05556 + 3 \cdot 5^2}{5^2\left(5-1\right)\left(5+1\right)} \\
&= \frac{25 \cdot 4{,}5 - 5 \cdot 21{,}05556 + 3 \cdot 25}{25 \cdot 4 \cdot 6} = \frac{112{,}5 - 105{,}2778 + 75}{600} \\
&= \frac{82{,}2222}{600} = 0{,}137037
\end{aligned}$$

και τελικά η διακύμανση του συντελεστή I υποθέτοντας κανονικότητα είναι:

$$Var(I) = E(I^2) - E(I)^2 = 0{,}137037 - (-0{,}25)^2 = 0{,}137037 - 0{,}0625 = 0{,}074537$$

Ο έλεγχος της μηδενικής υπόθεσης της απουσίας χωρικής αυτοσυσχέτισης γίνεται με τη στατιστική:

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}} \sim N(0,1)$$

η οποία για το συγκεκριμένο δείγμα θα είναι:

$$Z=\frac{I-E(I)}{\sqrt{Var(I)}}=\frac{0,41667-(-0,25)}{\sqrt{0,074537}}=\frac{0,6667}{0,2730147}=2,441871$$

Κατά συνέπεια για επίπεδο σημαντικότητας α = 5% η κριτική τιμή θα είναι $Z_{0,025}$ = 1,96 οπότε επειδή $|Z|=2,441871>1,96$ η μηδενική υπόθεση του ελέγχου απορρίπτεται και υπάρχει στατιστικά σημαντική θετική χωρική αυτοσυσχέτιση.

Β) Για να υπολογιστεί η διακύμανση της I υποθέτοντας τυχαιοποίηση πρέπει να χρησιμοποιηθεί για τον υπολογισμό της δεύτερη ροπή της η σχέση:

$$E(I^2)=\frac{n\left[\left(n^2-3n+3\right)S_1-nS_2+3S_0{}^2\right]-b_2\left[\left(n^2-n\right)S_1-2nS_2+6S_0{}^2\right]}{(n-1)(n-2)(n-3)S_0{}^2}$$

Ο συντελεστής κύρτωσης b_2 είναι:

$$b_2=\frac{\dfrac{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^4}{n}}{\left[\dfrac{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}{n}\right]^2}=\frac{\dfrac{2708}{5}}{\left[\dfrac{104}{5}\right]^2}=\frac{541,6}{20,8^2}=\frac{541,6}{432,64}=1,251849$$

Με αντικατάσταση του συντελεστή κύρτωσης b_2 και των ποσοτήτων S_0, S_1 και S_2 που έχουν ήδη υπολογιστεί όταν έγινε η υπόθεση της κανονικότητας στη σχέση της δεύτερης ροπής προκύπτει:

$$\begin{aligned}E(I^2)&=\frac{5\left[\left(5^2-3\cdot5+3\right)4,5-5\cdot21,05556+3\cdot5^2\right]-1,251849\left[\left(5^2-5\right)4,5-2\cdot5\cdot21,05556+6\cdot5^2\right]}{(5-1)(5-2)(5-3)5^2}\\&=\frac{5[13\cdot4,5-5\cdot21,05556+75]-1,251849[20\cdot4,5-10\cdot21,05556+150]}{4\cdot3\cdot2\cdot25}\\&=\frac{5\cdot28,2222-1,251849\cdot29,44444}{600}\\&=\frac{141,111-36,86}{600}=\frac{104,2511}{600}=0,1737518\end{aligned}$$

και τελικά η διακύμανση του συντελεστή I υποθέτοντας τυχαιοποίηση είναι:

$$Var(I)=E(I^2)-E(I)^2=0,1737518-(-0,25)^2=0,1737518-0,0625=0,1112518$$

Διαπιστώνεται ότι η διακύμανση του I με την υπόθεση της τυχαιοποίησης (0,112518) είναι μεγαλύτερη από τη διακύμανση που προέκυψε υποθέτοντας κανονικότητα (0,074537).

Όπως και στην υπόθεση της κανονικότητας ο έλεγχος της μηδενικής υπόθεσης της απουσίας χωρικής αυτοσυσχέτισης γίνεται με τη στατιστική:

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}} \sim N(0,1)$$

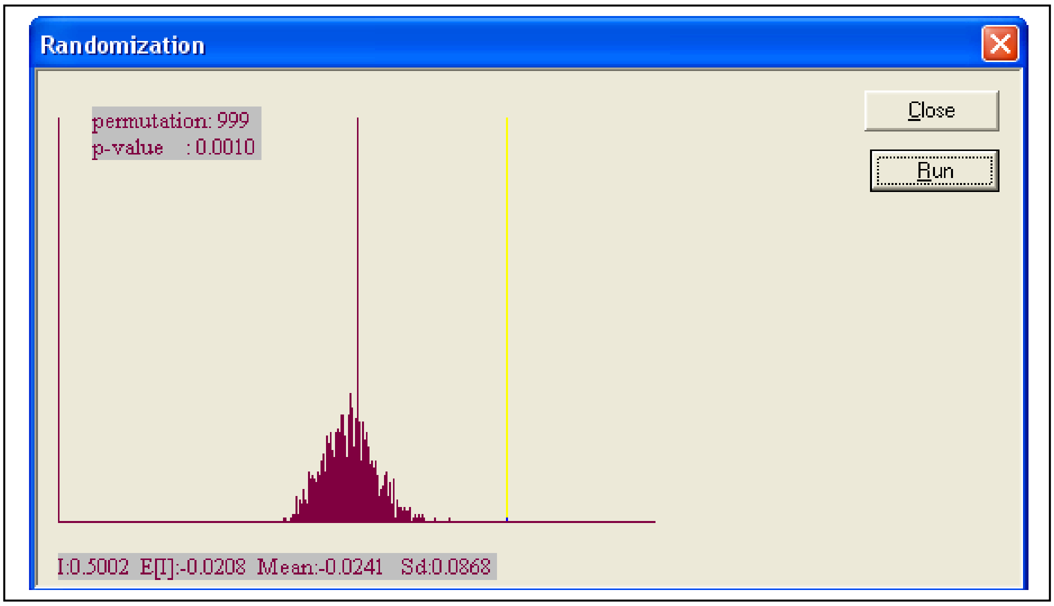
η οποία υπολογίζεται ως:

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}} = \frac{0,41667 - (-0,25)}{\sqrt{0,1112518}} = \frac{0,6667}{0,3335444} = 1,998735$$

Θεωρώντας και πάλι επίπεδο σημαντικότητας α = 5% η κριτική τιμή είναι $Z_{0,025}$ = 1,96 οπότε επειδή $|Z| = 1,998735 > 1,96$ η μηδενική υπόθεση του ελέγχου απορρίπτεται και με την υπόθεση της τυχαιοποίησης έστω και οριακά και υπάρχει στατιστικά σημαντική θετική χωρική αυτοσυσχέτιση.

Παράδειγμα 3.2

Με τη βοήθεια του λογισμικού χωρικής ανάλυσης GeoDa υπολογίστηκε ο συντελεστής I του Moran για τα συμβάντα εγκληματικότητας στην πρωτεύουσα του Οχάιο Columbus των ΗΠΑ που απεικονίζονται στο χάρτη στο Διάγραμμα 3.4 και κατασκευάστηκε χρησιμοποιώντας 999 μεταθέσεις η εμπειρική κατανομή του. Η μήτρα χωρικών σταθμίσεων ορίστηκε με εφαρμογή του κριτηρίου της βασίλισσας. Η οθόνη με τα αποτελέσματα του λογισμικού απεικονίζεται στο Διάγραμμα 3.6. Η τιμή του συντελεστή I που προέκυψε ισούται με 0,5002 που σημαίνει ότι υπάρχει ισχυρή θετική χωρική αυτοσυσχέτιση κάτι που συμπεράθηκε και από το χάρτη στο Διάγραμμα 3.4. Το λογισμικό μέσω της διαδικασίας των μεταθέσεων παρήγαγε μια εμπειρική κατανομή αναφοράς με την οποία συγκρίνεται η τιμή του I που έχει εκτιμηθεί από το δείγμα. Ταυτόχρονα, υπολόγισε μια τιμή πιθανότητας (*p-value*) για τον έλεγχο της μηδενικής υπόθεσης της απουσίας χωρικής αυτοσυσχέτισης. Η τιμή πιθανότητας ισούται με 0,0010 κάτι που συνεπάγεται ότι ο συντελεστής I είναι στατιστικά σημαντικός.



Διάγραμμα 3.6
Εμπειρική κατανομή της στατιστικής *I* για την εγκληματικότητα στο Columbus

3.4 Ο συντελεστής χωρικής αυτοσυσχετισης *I* του Moran με το υπόδειγμα γραμμικής παλινδρόμησης

Ο Anselin (1996) παρουσιάζει το συντελεστή I του Moran, για την περίπτωση που η μήτρα χωρικών σταθμίσεων είναι τυποποιημένη ανά γραμμή, σαν το συντελεστή της κλίσης μίας παλινδρόμησης με εξαρτημένη μεταβλητή τη χωρική υστέρηση και ανεξάρτητη μεταβλητή την εξεταζόμενη μεταβλητή σε αποκλίσεις από τη μέση τιμή της. Αυτό φαίνεται όταν υπολογιστεί ο I με τη βοήθεια άλγεβρας μητρών, δηλαδή:

$$I = \frac{(\mathbf{y}-\bar{\mathbf{y}})'\,\mathbf{W}(\mathbf{y}-\bar{\mathbf{y}})}{(\mathbf{y}-\bar{\mathbf{y}})'(\mathbf{y}-\bar{\mathbf{y}})} = \left[(\mathbf{y}-\bar{\mathbf{y}})'(\mathbf{y}-\bar{\mathbf{y}})\right]^{-1}(\mathbf{y}-\bar{\mathbf{y}})'\,\mathbf{W}(\mathbf{y}-\bar{\mathbf{y}}) = (\mathrm{y}'\mathrm{y})^{-1}\,\mathrm{y}'\mathbf{W}\mathrm{y}$$

όπου y είναι το $(n\times 1)$ διάνυσμα των τιμών της μεταβλητής y σε αποκλίσεις από τη μέση τιμή της και $\mathbf{W}\mathrm{y}$ το $(n\times 1)$ διάνυσμα της χωρικής υστέρησης. Όπως διαπιστώνεται ο συντελεστή I είναι ο εκτιμητής ελαχίστων τετραγώνων του γραμμικού υποδείγματος:

$$\mathbf{W}\mathrm{y} = \mathrm{y}I + \boldsymbol{\varepsilon}$$

Στην πραγματικότητα, ο συντελεστής I του Moran μπορεί να προκύψει εξίσου ως ο συντελεστής της κλίσης της παλινδρόμησης με εξαρτημένη μεταβλητή τη χωρική υστέρηση και ανεξάρτητη μεταβλητή την εξεταζόμενη μεταβλητή, δηλαδή χωρίς να γίνει ο μετασχηματισμός σε αποκλίσεις από τη μέση τιμή της μεταβλητής. Η μόνη διαφορά που υπάρχει αφορά το σταθερό όρο του υποδείγματος. Αυτό μπορεί να αποδειχθεί εύκολα με βάση τις ιδιότητες των εκτιμητών της παλινδρόμησης. Πάντως, επισημαίνεται ότι σε κάθε περίπτωση η διακύμανση του *I* που προκύπτει από την εκτίμηση του υποδείγματος παλινδρόμησης είναι λανθασμένη και αυτή θα πρέπει να υπολογιστεί με τη μεθοδολογία της προηγούμενης ενότητας.

Παράδειγμα 3.3

Θα υπολογιστεί με την εφαρμογή παλινδρόμησης ο συντελεστής *I* για τις πέντε χωρικές μονάδες στο Διάγραμμα 3.5 και τη μεταβλητή *y* του Πίνακα 3.1. Υπενθυμίζεται ότι η τυποποιημένη ανά γραμμή μήτρα χωρικών σταθμίσεων με το κριτήριο του πύργου είναι:

$$\mathbf{W} = \begin{bmatrix} 0{,}000\ 1{,}000\ 0{,}000\ 0{,}000\ 0{,}000 \\ 0{,}333\ 0{,}000\ 0{,}333\ 0{,}333\ 0{,}000 \\ 0{,}000\ 0{,}333\ 0{,}000\ 0{,}333\ 0{,}333 \\ 0{,}000\ 0{,}333\ 0{,}333\ 0{,}000\ 0{,}333 \\ 0{,}000\ 0{,}000\ 0{,}500\ 0{,}500\ 0{,}000 \end{bmatrix}$$

και το διάνυσμα των χωρικών υστερήσεων το:

$$\mathbf{Ly} = \mathbf{Wy} = \begin{bmatrix} 0{,}000\ 1{,}000\ 0{,}000\ 0{,}000\ 0{,}000 \\ 0{,}333\ 0{,}000\ 0{,}333\ 0{,}333\ 0{,}000 \\ 0{,}000\ 0{,}333\ 0{,}000\ 0{,}333\ 0{,}333 \\ 0{,}000\ 0{,}333\ 0{,}333\ 0{,}000\ 0{,}333 \\ 0{,}000\ 0{,}000\ 0{,}500\ 0{,}500\ 0{,}000 \end{bmatrix} * \begin{bmatrix} 5 \\ 6 \\ 16 \\ 14 \\ 14 \end{bmatrix} = \begin{bmatrix} 6{,}000 \\ 11{,}667 \\ 11{,}333 \\ 12{,}000 \\ 15{,}000 \end{bmatrix}$$

Ο συντελεστής *I* θα προκύψει από την εκτίμηση του γραμμικού υποδείγματος:

$$Ly_i = \alpha + \beta y_i + \varepsilon_i$$

Η διαδικασία υπολογισμού συνοψίζεται στον Πίνακα 3.2.

Πίνακας 3.2
Υπολογισμός συντελεστή *I* για τη μεταβλητή *y* με γραμμική παλινδρόμηση

Περιοχή i	Ly_i	y_i	$Ly_i \,.\, y_i$	y_i^2
1	6	5	30	25
2	11,667	6	70	36
3	11,333	16	181,333	256
4	12	14	168	196
5	15	14	210	196
Σύνολο	**56**	**55**	**659,333**	**709**

Η μέση τιμή της y είναι:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{5} y_i}{5} = \frac{5+6+16+14+14}{5} = \frac{55}{5} = 11$$

Η μέση τιμή της χωρικής υστέρησης είναι:

$$\overline{Ly} = \frac{\sum_{i=1}^{n} Ly_i}{n} = \frac{\sum_{i=1}^{5} Ly_i}{5} = \frac{6+11,667+11,333+12+15}{5} = \frac{56}{5} = 11,2$$

Ο εκτιμητής ελαχίστων τετραγώνων για το συντελεστή της κλίσης όπως είναι γνωστό δίνεται από τη σχέση:

$$\hat{\beta} = \frac{\sum_{i=1}^{5} y_i \cdot Ly_i - n \cdot \bar{y} \cdot \overline{Ly}}{\sum_{i=1}^{5} y_i^2 - n \cdot \bar{y}^2} = \frac{659,3333 - 5 \cdot 11 \cdot 11,2}{709 - 5 \cdot 11^2} = \frac{659,3333 - 616}{709 - 605} = \frac{43,3333}{104} = 0,41667$$

Ο εκτιμητής του σταθερού όρου του υποδείγματος θα είναι:

$$\hat{\alpha} = \overline{Ly} - \hat{\beta} \cdot \bar{y} = 11,2 - 11 \cdot 0,41667 = 11,2 - 4,584 = 6,62$$

Επομένως το εκτιμηθέν υπόδειγμα είναι:

$$\hat{L}y_i = 6,62 + 0,41667 y_i$$

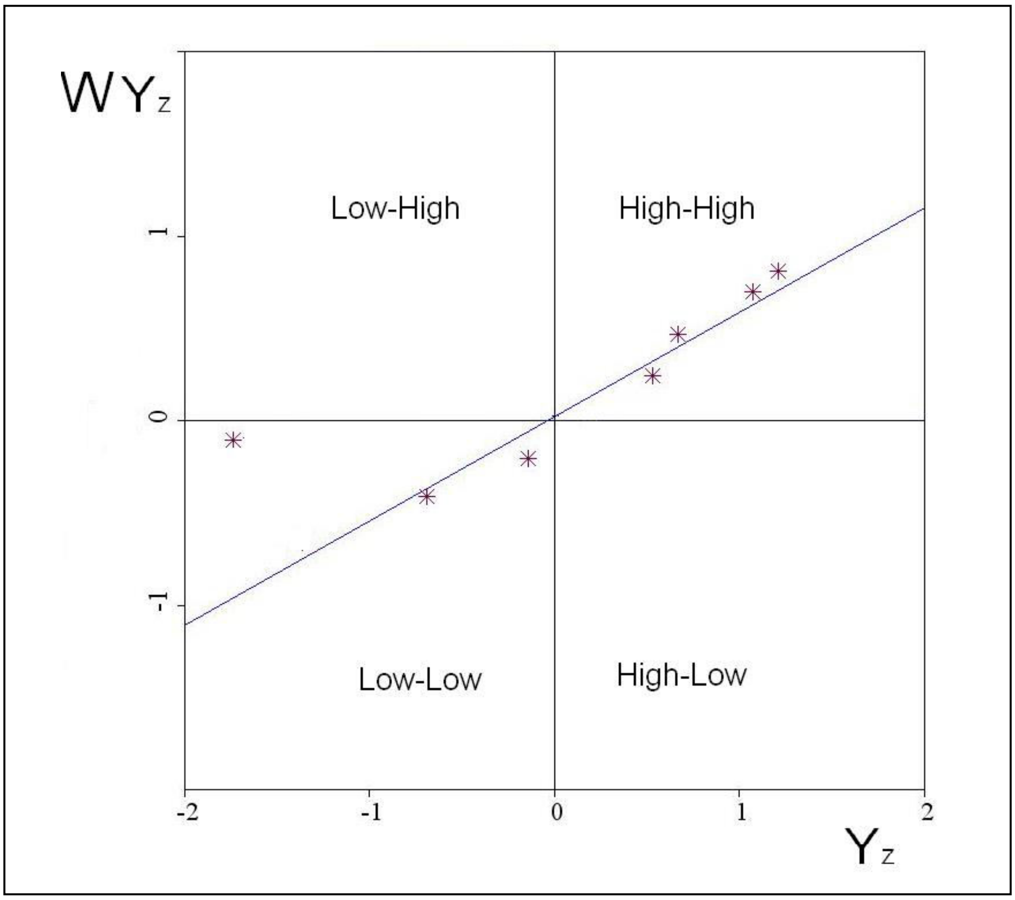
Κατά συνέπεια διαπιστώνεται ότι $\hat{\beta} = 0,41667 = I$ όπως αποδείχθηκε στο παράδειγμα 3.1.

3.5 Το διάγραμμα διασποράς του Moran

Το διάγραμμα διασποράς του Moran (*Moran scatter plot*) είναι ένα διάγραμμα διασποράς που στον οριζόντιο άξονα απεικονίζει μια μεταβλητή y στην τυποποιημένη της μορφή y_{zi} και στον κάθετο άξονα τη χωρική υστέρηση που προκύπτει για την τυποποιημένη μεταβλητή, δηλαδή:

$$y_{zi} = \frac{y_i - \bar{y}}{s_y} \qquad \text{και} \qquad Ly_{zi} = \sum_{j=1}^{n} w_{ij} y_{zj}$$

όπου $\bar{y}$ και s_y είναι αντίστοιχα η μέση τιμή και η τυπική απόκλιση της μεταβλητής *y*. Στο διάγραμμα διασποράς του Moran προστίθεται και η ευθεία παλινδρόμησης που της οποίας η κλίση ισούται με το συντελεστή *I*. Λόγω της τυποποίησης το διάγραμμα αυτό είναι κεντραρισμένο ως προς την αρχή των αξόνων. Ένα υποθετικό διάγραμμα διασποράς του Moran απεικονίζεται στο Διάγραμμα 3.7.



Διάγραμμα 3.7
Το διάγραμμα διασποράς του Moran

Ο τρόπος κατασκευής αυτού του διαγράμματος διασποράς επιτρέπει την εξέταση του τύπου της χωρικής αυτοσυσχέτισης που αντιστοιχεί σε κάθε χωρική μονάδα. Ειδικότερα, τα τέσσερα τεταρτημόρια που χωρίζεται το διάγραμμα αντιστοιχούν σε τέσσερεις διαφορετικούς τύπους χωρικής σχέσης μιας περιοχής με τις γειτονικές της περιοχές. Οι σχέσεις αυτές είναι:

1) Υψηλή-Υψηλή (High-High) Θετική χωρική αυτοσυσχέτιση

Στο τεταρτημόριο αυτό απεικονίζονται οι χωρικές μονάδες με υψηλές τιμές οι οποίες έχουν γείτονες με επίσης υψηλές τιμές. Δηλαδή πρόκειται για χωρικές συστάδες (*spatial clusters*). Ο τύπος αυτής της σχέσης για μια χωρική μονάδα πολλές φορές στη χωρική ανάλυση ονομάζεται καυτό σημείο (*hot spot*).

2) Χαμηλή-Χαμηλή (Low-Low) Θετική χωρική αυτοσυσχέτιση

Στο τεταρτημόριο αυτό απεικονίζονται οι χωρικές μονάδες με χαμηλές τιμές οι οποίες έχουν γείτονες με επίσης χαμηλές τιμές. Δηλαδή πρόκειται για χωρικές συστάδες (*spatial clusters*). Ο τύπος αυτής της σχέσης για μια χωρική μονάδα πολλές φορές στη χωρική ανάλυση ονομάζεται κρύο σημείο (*cold spot*).

3) Χαμηλή-Υψηλή (Low-High) Αρνητική χωρική αυτοσυσχέτιση

Στο τεταρτημόριο αυτό απεικονίζονται οι χωρικές μονάδες με χαμηλες τιμές οι οποίες έχουν γείτονες με υψηλές τιμές.

4) Υψηλή-Χαμηλή (High-Low) Αρνητική χωρική αυτοσυσχέτιση

Στο τεταρτημόριο αυτό απεικονίζονται οι χωρικές μονάδες με υψηλές τιμές οι οποίες έχουν γείτονες με χαμηλές τιμές.

Εφόσον ο δείκτης I έχει θετική τιμή οι περιοχές στα τεταρτημόρια Χαμηλή-Υψηλή (Low-High) και Υψηλή-Χαμηλή (High-Low) θα είναι χωρικά έκτροπες παρατηρήσεις (*spatial outliers).*

Παράδειγμα 3.4

Για να γίνουν κατανοητοί ο τρόπος κατασκευής και η ερμηνεία του διαγράμματος διασποράς του Moran θα παρουσιαστεί αναλυτικά η δημιουργία του για τη μεταβλητή y του Πίνακα 3.1. Αρχικά πρέπει να υπολογιστεί η μέση τιμή και η τυπική απόκλιση της μεταβλητής για να γίνει η τυποποίηση. Η μέση τιμή της μεταβλητής y όπως έχει δειχτεί σε προηγούμενο παράδειγμα είναι $\bar{y} = 11$. Η τυπική της απόκλιση θα είναι:

$$s_y = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{5}\left(y_i - 11\right)^2}{5-1}} = \sqrt{\frac{104}{4}} = \sqrt{26} = 5{,}1$$

Η τυποποίηση της μεταβλητής θα γίνει με την εφαρμογή της σχέσης:

$$y_{zi} = \frac{y_i - \bar{y}}{s_y} = \frac{y_i - 11}{5{,}1}$$

Επομένως, το (5×1) διάνυσμα των τιμών της τυποποιημένης μεταβλητής y_{Zi} θα είναι:

$$\mathbf{y}_Z = \begin{bmatrix} -1{,}1767 \\ -0{,}9806 \\ 0{,}9806 \\ 0{,}5883 \\ 0{,}5883 \end{bmatrix}$$

Στη συνέχεια υπολογίζεται η χωρική υστέρηση για την τυποποιημένη μεταβλητή ως εξής:

$$\mathbf{Ly}_Z = \mathbf{Wy}_Z = \begin{bmatrix} 0{,}000 & 1{,}000 & 0{,}000 & 0{,}000 & 0{,}000 \\ 0{,}333 & 0{,}000 & 0{,}333 & 0{,}333 & 0{,}000 \\ 0{,}000 & 0{,}333 & 0{,}000 & 0{,}333 & 0{,}333 \\ 0{,}000 & 0{,}333 & 0{,}333 & 0{,}000 & 0{,}333 \\ 0{,}000 & 0{,}000 & 0{,}500 & 0{,}500 & 0{,}000 \end{bmatrix} * \begin{bmatrix} -1{,}1767 \\ -0{,}9806 \\ 0{,}9806 \\ 0{,}5883 \\ 0{,}5883 \end{bmatrix} = \begin{bmatrix} -0{,}9806 \\ 0{,}1307 \\ 0{,}0654 \\ 0{,}1961 \\ 0{,}7845 \end{bmatrix}$$

Πίνακας 3.3
Τιμές τυποποιημένης μεταβλητής και χωρικές υστερήσεις

Περιοχή i	y_i	y_{zi}	L_{yzi}
1	5	-1,1767	-0,9806
2	6	-0,9806	0,1307
3	16	0,9806	0,0654
4	14	0,5883	0,1961
5	14	0,5883	0,7845

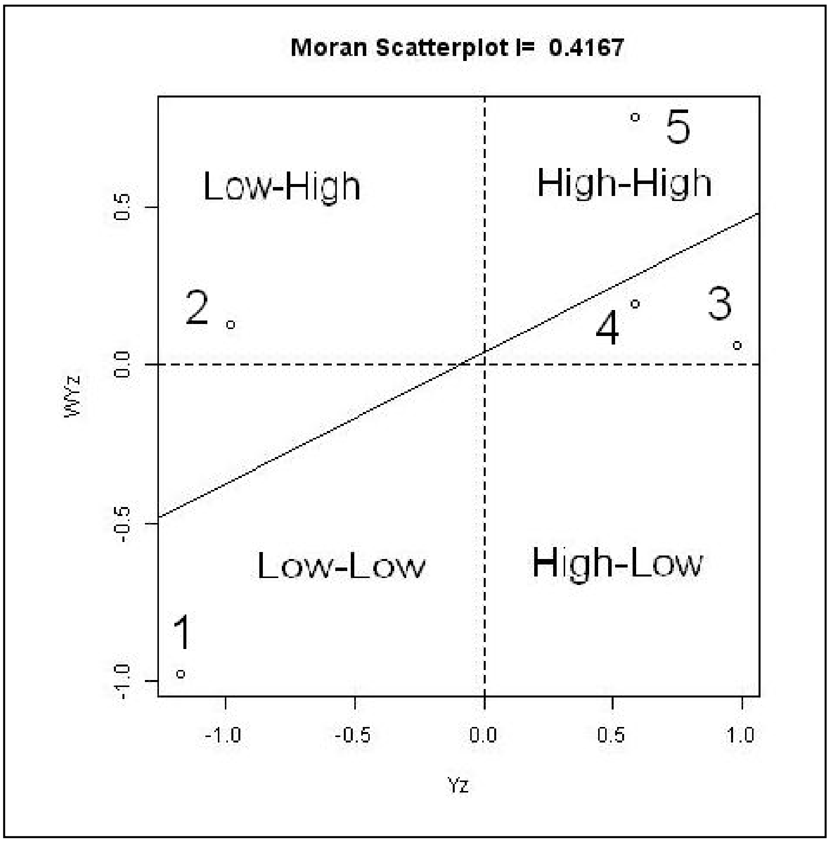
Στον Πίνακα 3.3 συνοψίζονται τα αποτελέσματα. Οι τιμές y_{zi} και L_{yzi} θα απεικονιστούν ως σημεία στο διάγραμμα διασποράς. Το επόμενα στάδιο που πρέπει να γίνει είναι η εκτίμηση της ευθείας παλινδρόμησης:

$$Ly_{zi} = \alpha + \beta yz_i + \varepsilon_i$$

η οποία θα συμπεριληφθεί στο διάγραμμα. Η εκτιμημένη ευθεία είναι η:

$$\hat{L}y_{zi} = 0,03922 + 0,41667\,yz_i$$

και προκύπτει με παρόμοιο τρόπο με αυτόν που παρουσιάστηκε αναλυτικά στο Παράδειγμα 3.3. Το τελικό διάγραμμα διασποράς του Moran απεικονίζεται στο Διάγραμμα 3.3.



Διάγραμμα 3.8
Το διάγραμμα διασποράς του Moran για τη μεταβλητή του Πίνακα 3.3

Όπως φαίνεται στο Διάγραμμα 3.8 οι χωρικές μονάδες 3, 4 και 5 ανήκουν στο τεταρτημόριο High-High. Δηλαδή είναι περιοχές με υψηλές τιμές που γειτονεύουν με περιοχές με επίσης υψηλές τιμές (*hot-spot*). Η χωρική μονάδα 1 ανήκει στο τεταρτημόριο Low-Low και είναι περιοχή χαμηλών τιμών που έχει γείτονες με επίσης χαμηλές τιμές (*cold spot*). Η χωρική μονάδα 2 είναι χωρικά έκτροπη παρατήρηση (*spatial outlier)* καθώς ανήκει στο τεταρτημόριο Low-High. Δηλαδή είναι περιοχή χαμηλών τιμών που έχει γείτονες με υψηλές τιμές παρουσιάζοντας αρνητική χωρική αυτοσυσχέτιση.

3.6 Οι τοπικοί συντελεστές I_i του Moran

Οι τοπικοί συντελεστές I_i του Moran ανήκουν σε μια ευρύτερη κατηγορία στατιστικών μέτρων που ονομάζονται τοπικές στατιστικές συναρτήσεις (*local statistics*). Οι συναρτήσεις αυτής της μορφής έχουν σκοπό να μελετήσουν ένα χαρακτηριστικό σε μια συγκεκριμένη περιοχή του γεωγραφικού χώρου περιλαμβάνοντας στον υπολογισμό τους τη σχέση που υπάρχει με τις γειτονικές περιοχές. Η βασική τους ιδιότητα είναι ότι έχουν πολυμεταβλητή διάσταση καθώς ο υπολογισμός τους καταλήγει στον προσδιορισμό διαφορετικής αριθμητικής τιμής για κάθε χωρική μονάδα. Ως συνέπεια, για να αξιοποιηθούν από τον ερευνητή απαιτείται η χαρτογράφηση των τιμών τους η οποία διευκολύνει στον εντοπισμό των χωρικών διαφοροποιήσεων της εξεταζόμενης μεταβλητής στην υπό μελέτη γεωγραφική περιοχή.

Οι τοπικοί (*local*) δείκτες χωρικής αυτοσυσχέτισης προτάθηκαν από τον Anselin (1995) και προκύπτουν από τη διάσπαση των ολικών (*global*) δεικτών χωρικής αυτοσυσχέτισης σε επιμέρους δείκτες που ο Anselin (1995) τους ονόμασε τοπικούς δείκτες χωρικής σχέσης (*Local Indicators of Spatial Association*-LISA). Ειδικότερα, ο τοπικός δείκτης I_i του Moran μετράει την ύπαρξη χωρικής αυτοσυσχέτισης γύρω από μία παρατήρηση (χωρική μονάδα). Σε αντίθεση με τον ολικό δείκτη I του Moran που παρουσιάστηκε στις προηγούμενες ενότητες αυτού του κεφαλαίου ο τοπικός δείκτης αποβλέπει στο να εντοπίσει την ύπαρξη χωρικής ετερογένειας στην εξεταζόμενη περιοχή. Ο τοπικός δείκτης I_i του Moran όταν η μήτρα χωρικών σταθμίσεων είναι τυποποιημένη ανά γραμμή ορίζεται ως εξής:

$$I_i = \frac{\left(y_i - \bar{y}\right)\sum_{j=1}^{n} w_{ij}\left(y_j - \bar{y}\right)}{\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}{n}}$$

Η αναμενόμενη τιμή του τοπικού δείκτη I_i είναι: $E(I_i) = \frac{-1}{n-1}$

Η ερμηνεία του τοπικού δείκτη I_i είναι ουσιαστικά η ίδια με αυτή του ολικού δείκτη I, κατά συνέπεια:

- Εάν $I_i > E(I_i)$, τότε υπάρχει θετική τοπική χωρική αυτοσυσχέτιση, δηλαδή οι γειτονικές χωρικές μονάδες της *i* χωρικής μονάδας έχουν παρόμοιες υψηλές η χαμηλές τιμές.
- Όταν $I_i < E(I_i)$ τότε υπάρχει αρνητική τοπική χωρική αυτοσυσχέτιση και οι γειτονικές χωρικές μονάδες της *i* χωρικής μονάδας έχουν διαφορετικές σε μέγεθος τιμές από αυτή.

Οι τοπικοί δείκτες I_i μπορούν να δείξουν τη συνεισφορά κάθε χωρικής μονάδας ξεχωριστά στον υπολογισμό της τιμής του ολικού δείκτη *I* και να βοηθήσουν στον εντοπισμό τοπικών χωρικών συστάδων (*spatial clusters*) και τοπικών χωρικών έκτροπων παρατηρήσεων (*spatial outliers*) επικουρώντας στην ανάλυση με το διάγραμμα διασποράς του Moran. Αυτό οφείλεται στην ιδιότητα των τοπικών δεικτών ότι:

$$\frac{\sum_{i=1}^{n} I_i}{n} = I$$

δηλαδή, ο μέσος όρος τους για όλη τη γεωγραφική περιοχή ισούται με την τιμή του ολικού δείκτη *I*.

Για να εντοπιστούν οι χωρικές διαφοροποιήσεις στην εξεταζόμενη γεωγραφική περιοχή οι τιμές των τοπικών δεικτών I_i συγκρίνονται με την τιμή του ολικού δείκτη. Πιο συγκεκριμένα, όταν το πρόσημο ενός τοπικού δείκτη είναι διαφορετικό από αυτό του ολικού δείκτη η παρατήρηση στην οποία αντιστοιχεί ο τοπικός δείκτης θα είναι έκτροπη (*spatial outlier*). Επιπρόσθετα, εάν η τιμή ενός τοπικού δείκτη είναι μεγαλύτερη από την τιμή του ολικού δείκτη τότε υπάρχει χωρική συγκέντρωση παρόμοιων τιμών (υψηλών ή χαμηλών) γύρω από τη συγκεκριμένη παρατήρηση. Δηλαδή, εντοπίζεται μια χωρική συστάδα (*spatial cluster*). Αντίθετα, εάν η τιμή ενός τοπικού δείκτη είναι μικρότερη από την τιμή του ολικού δείκτη τότε δεν θα υπάρχει χωρική συγκέντρωση παρόμοιων τιμών (υψηλών ή χαμηλών) γύρω από τη συγκεκριμένη παρατήρηση.

Παράδειγμα 3.5

Θα υπολογιστούν οι τοπικοί δείκτες I_i για τις πέντε χωρικές μονάδες στο Διάγραμμα 3.1 με τη μεταβλητή *y* του Πίνακα 3.1.

Η μέση τιμή της y είναι:

$$\bar{y}=\frac{\sum_{i=1}^{n} y_i}{n}=\frac{\sum_{i=1}^{5} y_i}{5}=\frac{5+6+16+14+14}{5}=\frac{55}{5}=11$$

Ο μέσος όρος του αθροίσματος των τετραγώνων των αποκλίσεων των τιμών της y από τη μέση τιμή της είναι:

$$\frac{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}{n}=\frac{\sum_{i=1}^{5}\left(y_i-11\right)^2}{5}=\frac{104}{5}=20,8$$

Η τυποποιημένη ανά γραμμή μήτρα χωρικών σταθμίσεων με το κριτήριο του πύργου είναι:

$$\mathbf{W}=\begin{bmatrix} 0,000 & 1,000 & 0,000 & 0,000 & 0,000 \\ 0,333 & 0,000 & 0,333 & 0,333 & 0,000 \\ 0,000 & 0,333 & 0,000 & 0,333 & 0,333 \\ 0,000 & 0,333 & 0,333 & 0,000 & 0,333 \\ 0,000 & 0,000 & 0,500 & 0,500 & 0,000 \end{bmatrix}$$

Ο τοπικός δείκτης I_i για κάθε χωρική μονάδα θα προκύψει από την εφαρμογή της σχέσης:

$$I_i=\frac{\left(y_i-\bar{y}\right)\sum_{j=1}^{5} w_{ij}\left(y_j-\bar{y}\right)}{\frac{\sum_{i=1}^{5}\left(y_i-\bar{y}\right)^2}{n}}=\frac{\left(y_i-11\right)}{20,8}\sum_{j=1}^{5} w_{ij}\left(y_j-11\right)$$

Οι τιμές των τοπικών δεικτών που προκύπτουν περιλαμβάνονται στον Πίνακα 3.4 μαζί με τις τιμές της μεταβλητής y και την τιμή του ολικού δείκτη (I = 0,41667) που υπολογίστηκε στο Παράδειγμα 3.1. Αναλυτικά οι τοπικοί δείκτες για κάθε χωρική μονάδα προκύπτουν ως εξής:

$$I_1=\frac{(5-11)}{20,8}\sum_{j=1}^{5} w_{ij}\left(y_j-11\right)=\frac{(5-11)}{20,8}\cdot\left[1\cdot(6-11)\right]=-0,288\cdot(-5)=1,442$$

$$I_2=\frac{(6-11)}{20,8}\sum_{j=1}^{5} w_{ij}\left(y_j-11\right)=\frac{(6-11)}{20,8}\cdot\left[0,333\cdot(5-11)+0,333\cdot(16-11)+0,333\cdot(14-11)\right]$$
$$=-0,240\cdot 0,6667=-0,160$$

$$I_3=\frac{(16-11)}{20,8}\sum_{j=1}^{5} w_{ij}\left(y_j-11\right)=\frac{(16-11)}{20,8}\cdot\left[0,333\cdot(6-11)+0,333\cdot(14-11)+0,333\cdot(14-11)\right]$$
$$=0,240\cdot 0,3333=0,0801$$

$$I_4 = \frac{(14-11)}{20,8}\sum_{j=1}^{5} w_{ij}\left(y_j - 11\right) = \frac{(14-11)}{20,8}\cdot\left[0,333\cdot(6-11)+0,333\cdot(16-11)+0,333\cdot(14-11)\right]$$

$$= 0,144\cdot 1 = 0,144$$

$$I_4 = \frac{(14-11)}{20,8}\sum_{j=1}^{5} w_{ij}\left(y_j - 11\right) = \frac{(14-11)}{20,8}\cdot\left[0,5\cdot(16-11)+0,5\cdot(14-11)\right]$$

$$= 0,144\cdot 4 = 0,577$$

Η αναμενόμενη τιμή για κάθε τοπικό δείκτη είναι:

$$E(I_i) = \frac{-1}{n-1} = \frac{-1}{5-1} = \frac{-1}{4} = -0,25$$

Πίνακας 3.4
Τοπικοί δείκτες I_i του Moran

Περιοχή i	y_i	Local I_i	$E(I_i)$	Global I
1	5	1,442	-0,25	0,41667
2	6	-0,160	-0,25	0,41667
3	16	0,0801	-0,25	0,41667
4	14	0,144	-0,25	0,41667
5	14	0,577	-0,25	0,41667

Τα βασικά συμπεράσματα που προκύπτουν είναι:

- Για τη χωρική μονάδα 1 ισχύει I_1 = 1,442 > $E(I_i)$ = -0,25 δηλαδή υπάρχει θετική τοπική χωρική αυτοσυσχέτιση. Επειδή η τιμή του τοπικού δείκτη είναι κατά πολύ μεγαλύτερη από την τιμή του ολικού δείκτη (I = 0,41667) υπάρχει χωρική συγκέντρωση γύρω από τη χωρική μονάδα 1.
- Για τη χωρική μονάδα 2 ισχύει I_2 = -0,160> $E(I_i)$ = -0,25 δηλαδή υπάρχει θετική τοπική χωρική αυτοσυσχέτιση. Επειδή η τιμή του τοπικού δείκτη έχει αντίθετο πρόσημο από την τιμή του ολικού δείκτη (I = 0,41667) πρόκειται για έκτροπη παρατήρηση (*spatial outlier*).
- Για τη χωρική μονάδα 3 ισχύει I_3 = 0,0801 > $E(I_i)$ = -0,25 δηλαδή υπάρχει θετική τοπική χωρική αυτοσυσχέτιση. Επειδή η τιμή του τοπικού δείκτη είναι μικρότερη από την τιμή του ολικού δείκτη (I = 0,41667) δεν υπάρχει χωρική συγκέντρωση γύρω από τη χωρική μονάδα 3.

- Για τη χωρική μονάδα 4 ισχύει I_4 = 0,144 > $E(I_i)$ = -0,25 δηλαδή υπάρχει θετική τοπική χωρική αυτοσυσχέτιση. Επειδή η τιμή του τοπικού δείκτη είναι μικρότερη από την τιμή του ολικού δείκτη (Ι = 0,41667) δεν υπάρχει χωρική συγκέντρωση γύρω από τη χωρική μονάδα 4.
- Για τη χωρική μονάδα 5 ισχύει I_1 = 0,577 > $E(I_i)$ = -0,25 δηλαδή υπάρχει θετική τοπική χωρική αυτοσυσχέτιση. Επειδή η τιμή του τοπικού δείκτη είναι μεγαλύτερη από την τιμή του ολικού δείκτη (Ι = 0,41667) υπάρχει χωρική συγκέντρωση γύρω από τη χωρική μονάδα 5.

Τέλος, διαπιστώνεται ότι ισχύει η ιδιότητα ότι ο μέσος όρος όλων των τοπικών δεικτών ισούται με την τιμή του ολικού δείκτη καθώς προκύπτει:

$$\frac{\sum_{i=1}^{5} I_i}{5} = \frac{1,442 - 0,160 + 0,0801 + 0,144 + 0,577}{5} = \frac{2,0831}{5} = 0,41667 = I$$

3.7 Ο συντελεστής *C* του Geary

Ο συντελεστής *C* του Geary (1954) συγκρίνει τις τιμές μιας μεταβλητής *y* με τις τιμές της σε γειτονικές περιοχές για να εκτιμήσει την ένταση της χωρικής αυτοσυσχέτισης. Η αλγεβρική του μορφή είναι η εξής:

$$C = \frac{(n-1)}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(y_i - y_j\right)}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}$$

Όπως και στον συντελεστή *I* του Moran όταν η μήτρα χωρικών σταθμίσεων είναι τυποποιημένη ανά γραμμή θα ισχύει $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} = n$ οπότε ο δείκτης θα γράφεται:

$$C = \frac{(n-1)}{2n} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(y_i - y_j\right)}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}$$

Όταν οι χωρικές μονάδες έχουν παρόμοιες τιμές (θετική χωρική αυτοσυσχέτιση) ο αριθμητής στο προηγούμενο πηλίκο και κατά συνέπεια και ο δείκτης θα είναι κοντά στο μηδέν καθώς για κάθε ζευγάρι παρατηρήσεων η διαφορά (y_i-y_j) θα είναι κοντά στο μηδέν. Αντίθετα, όταν οι χωρικές μονάδες έχουν ανόμοιες τιμές (αρνητική χωρική αυτοσυσχέτιση) ο δείκτης θα πλησιάζει στο δύο. Ο συντελεστής C λαμβάνει τιμές στο διάστημα [0, 2] και η αναμενόμενη τιμή του όταν δεν υπάρχει χωρική αυτοσυσχέτιση είναι $E(C)= 1$. Ειδικότερα, ως προς την ερμηνεία του δείκτη ισχύουν τα εξής:

- Εάν $C < E(C) = 1$ υπάρχει θετική χωρική αυτοσυσχέτιση.
- Εάν $C > E(C) = 1$ υπάρχει αρνητική χωρική αυτοσυσχέτιση.
- Όταν $C = E(C) = 1$ δεν υπάρχει χωρική αυτοσυσχέτιση.

Ο έλεγχος για τη στατιστική σημαντικότητα του συντελεστή C διεξάγεται με παρόμοια μεθοδολογία όπως και για το συντελεστή I του Moran. Δηλαδή, υπολογίζεται η διακύμανσή του υποθέτοντας κανονικότητα η τυχαιοποίηση και στη συνέχεια γίνεται ο μετασχηματισμός σε τυποποιημένη κανονική κατανομή. Οι αλγεβρικοί τύποι για των υπολογισμό των διακυμάνσεων υπάρχουν στους Cliff και Ord (1981).

3.8 Ο δείκτης *G* των Getis και Ord

Ο δείκτης G έχει προταθεί από τους Getis και Ord (1992) και εξετάζει τη συνολική συγκέντρωση ή την έλλειψη συγκέντρωσης σε ζευγάρια γειτονικών τιμών. Σκοπός του είναι να διερευνήσει, στην περίπτωση που οι δείκτες I και C έχουν υποδείξει την ύπαρξη θετικής χωρικής αυτοσυσχέτισης, εάν υπάρχουν συστάδες υψηλών ή χαμηλών τιμών. Δηλαδή, χρησιμοποιείται αποκλειστικά όταν υπάρχει θετική χωρική αυτοσυσχέτιση. Ο δείκτης υπολογίζεται ως εξής:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} y_i y_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j} \qquad \text{για} \quad j \neq i$$

Η αναμενόμενη τιμή του δείκτη είναι:

$$E(G) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}{n(n-1)} \qquad \text{για} \quad j \neq i$$

Η ερμηνεία του δείκτη είναι:

- Εάν $G > E(G)$ τότε υπάρχουν συστάδες υψηλών τιμών.
- Εάν $G < E(G)$ τότε υπάρχουν συστάδες χαμηλών τιμών.

Ο δείκτης λαμβάνει τιμές στο διάστημα [0, 1]. Επιπρόσθετα, πρέπει να σημειωθεί ότι οι παρατηρήσεις της εξεταζόμενης μεταβλητής πρέπει απαραίτητα να έχουν θετικές τιμές. Επίσης, ο δείκτης αδυνατεί να διακρίνει την ύπαρξη αρνητικής χωρικής αυτοσυσχέτισης. Στην περίπτωση που υπάρχει αρνητική χωρική αυτοσυσχέτιση και εσφαλμένα υπολογιστεί ο δείκτης τότε θα λάβει μικρή τιμή και θα παραπέμψει λανθασμένα στην ύπαρξη συστάδων χαμηλών τιμών.