ΚΕΦΑΛΑΙΟ 1

ΧΩΡΙΚΑ ΔΕΔΟΜΕΝΑ

1.1 Εισαγωγή

Η ραγδαία βελτίωση της ισχύς των ηλεκτρονικών υπολογιστών που έχει επιτευχθεί τα τελευταία χρόνια και η ανάπτυξη των γεωγραφικών συστημάτων πληροφοριών έχουν συμβάλλει στην εμφάνιση και στην ανάγκη για μελέτη μιας νέας μορφής στατιστικών δεδομένων στην οποία τα παλιότερα χρόνια δεν είχε δοθεί η δέουσα προσοχή. Ο νέος τύπος δεδομένων αφορά τα χωρικά δεδομένα (*spatial data*) δηλαδή, διαστρωματικά δεδομένα που έχουν συλλεχθεί και προέρχονται από διαφορετικές γεωγραφικές περιοχές.

Η διαφορά των χωρικών δεδομένων από τα υπόλοιπα διαστρωματικά δεδομένα βρίσκεται στο γεγονός πώς κάθε παρατήρηση του δείγματος συνοδεύεται από γεωγραφική αναφορά ως προς τη γεωγραφική περιοχή που έχει συλλεχθεί. Πιο συγκεκριμένα, η παρατήρηση μπορεί να καταγράφεται μαζί με τις γεωγραφικές συντεταγμένες του σημείου του χώρου στο οποίο αναφέρεται ή να ανήκει σε κάποιο πολύγωνο (π.χ. νομός, χώρα) ενός χάρτη.

1.2 Βασικές κατηγορίες χωρικών δεδομένων

Στη βιβλιογραφία εμφανίζονται αρκετές διαφορετικές προσεγγίσεις με τις οποίες μπορούν να κατηγοριοποιηθούν οι διαφορετικοί τύποι των χωρικών δεδομένων. Η ταξινόμηση που ακολουθεί βασίζεται στη μεθοδολογία των Cressie (1993) και Schabenberger και Gotway (2005) σύμφωνα με την οποία τα χωρικά δεδομένα προέρχονται από την πραγματοποίηση κάποιας χωρικής στοχαστικής διαδικασίας και η διάκρισή τους γίνεται με βάση τα χαρακτηριστικά της χωρικής περιοχής *D*. Πιο συγκεκριμένα, έστω μια χωρική διαδικασία:

$$\left\{ Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathrm{R}^d \right\}$$

όπου

- Z είναι το χαρακτηριστικό που μελετάται στο χώρο (παραδείγματος χάριν θάνατοι, γεννήσεις, θερμοκρασία, ρύπανση, ανεργία, εισόδημα).

- **s** είναι ένα ($d \times 1$) διάνυσμα που περιλαμβάνει τις χωρικές συντεταγμένες της τοποθεσίας που παρατηρείται η *Z*.
- *D* είναι ένα σύνολο που δηλώνει τη χωρική περιοχή (*spatial domain*).

Συνήθως στις περισσότερες περιπτώσεις ισχύει ότι $d = 2$ διότι γίνεται αναφορά σε γεωγραφικές συντεταγμένες στο επίπεδο οπότε το διάνυσμα **s** θα είναι διαστάσεων (2×1) και θα περιέχει Καρτεσιανές συντεταγμένες (*Cartesian coordinates*).

Οι δείκτες στο σύνολο *D* μπορεί να είναι:

α) Συνεχείς όταν αναφέρονται σε γεωγραφικές συντεταγμένες σημείων στο χώρο.

β) Διακριτοί όταν αναφέρονται σε συντεταγμένες σημείων επάνω σε ένα ομαλό πλέγμα ή σε αύξοντες αριθμούς γεωγραφικών περιοχών.

Επομένως, σύμφωνα με τα χαρακτηριστικά της χωρικής περιοχής *D* τα χωρικά δεδομένα χωρίζονται σε τρείς βασικές κατηγορίες.

1) Γεωστατιστικά δεδομένα (*geostatistical data*) ή δεδομένα σημείου (*point data*)

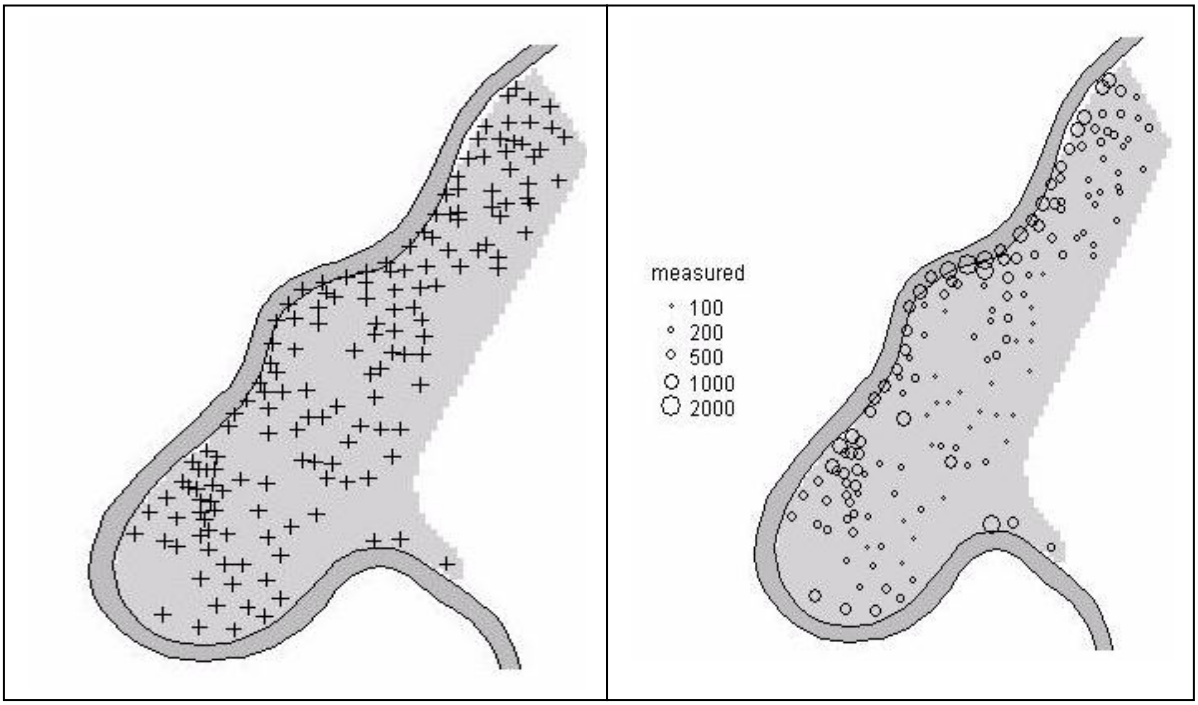
Στα γεωστατιστικά δεδομένα η χωρική περιοχή *D* είναι ένα συνεχές σταθερό σύνολο (δηλαδή τα σημεία που ορίζουν τη χωρική περιοχή *D* είναι μη στοχαστικά) που σημαίνει ότι η $Z(\mathbf{s})$ μπορεί να παρατηρηθεί οπουδήποτε μέσα στο χώρο που καθορίζεται από τη χωρική περιοχή *D*. Εξ αιτίας της συνέχειας της χωρικής περιοχής *D* τα γεωστατιστικά δεδομένα συχνά αποκαλούνται και ως χωρικά δεδομένα συνεχούς διακύμανσης (*spatial data with continuous variation*). Επισημαίνεται ότι η συνέχεια αφορά τη χωρική περιοχή και όχι το χαρακτηριστικό που παρατηρείται το οποίο μπορεί να διακριτό η συνεχές. Επειδή η χωρική περιοχή είναι συνεχής δεν είναι δυνατόν να μελετηθεί εξονυχιστικά και για το λόγο αυτό ο ερευνητής συνήθως ενδιαφέρεται να εξετάσει τη γεωγραφική μεταβλητότητα του χαρακτηριστικού με τεχνικές όπως η χαρτογράφηση.

2) Δικτυωτά δεδομένα (*lattice data*) ή Επιφανειακά ή Περιφερειακά δεδομένα (*areal* ή *regional data*).

Εδώ η χωρική περιοχή *D* είναι ένα διακριτό σταθερό σύνολο δηλαδή μη τυχαίο και μετρήσιμο. Οι περιοχές μπορεί να είναι άπειρες αλλά μετρήσιμες. Σε αυτήν την κατηγορία, τα δεδομένα ομαδοποιούνται και θεωρούνται ότι ανήκουν σε ένα πολύγωνο ενός χάρτη όπως είναι οι χώρες, οι νομοί, οι περιφέρειες ή ακόμη και τα απογραφικά τετράγωνα. Εάν υπάρχει η ανάγκη για αναφορά με γεωγραφικές συντεταγμένες τότε συνήθως χρησιμοποιούνται είτε οι συντεταγμένες κάποιας πόλης όπως η πρωτεύουσα της χώρας ή του νομού είτε, τις περισσότερες φορές, το γεωγραφικό κεντροειδές (*geographic centroid*) του πολυγώνου. Με

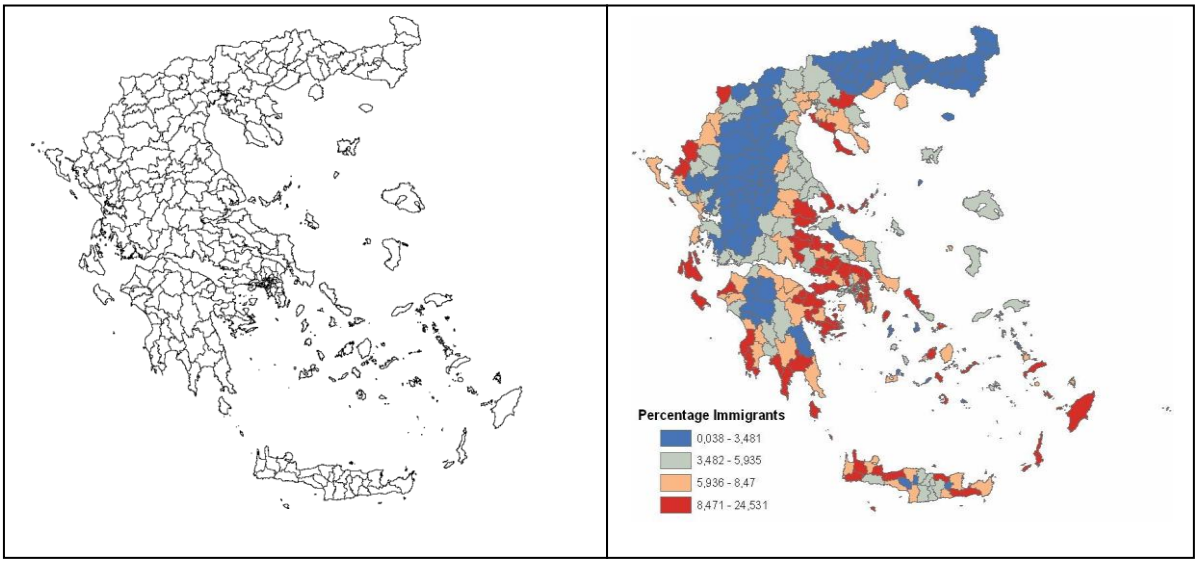
τα περιφερειακά δεδομένα συνήθως ο ερευνητής κατασκευάζει υποδείγματα και μελετάει τις στατιστικές σχέσεις μεταξύ διαφόρων μεταβλητών.

Για να γίνει κατανοητή η διάκριση μεταξύ των δύο κατηγόριων δεδομένων θα μπορούσαν να αναφερθούν ως παραδείγματα των γεωστατιστικών δεδομένων το ύψος της βροχόπτωσης, το επίπεδο της ατμοσφαιρικής ρύπανσης και η μόλυνση στο έδαφος από κάποια ουσία. Μια τέτοια επιστημονική μέτρηση θεωρητικά θα μπορούσε να γίνει από τον ερευνητή με το κατάλληλο επιστημονικό όργανο σε οποιοδήποτε γεωγραφικό σημείο γνωρίζοντας ταυτόχρονα και τις γεωγραφικές του συντεταγμένες. Αντίστοιχα, για τα περιφερειακά δεδομένα παραδείγματα αποτελούν το ύψος της ανεργίας, το εισόδημα, ο αριθμός των γεννήσεων και ο αριθμός των αλλοδαπών. Τα δεδομένα αυτά συλλέγονται μέσα σε μία γεωγραφική περιοχή, αθροίζονται και αποδίδονται σε αυτή, χωρίς να υπάρχει η δυνατότητα ταυτοποίησης γεωγραφικών συντεταγμένων για κάθε μεμονωμένη παρατήρηση. Κατά συνέπεια, θα μπορούσε να θεωρηθεί ότι στα γεωστατιστικά δεδομένα ο γεωγραφικός χώρος υποτίθεται ότι είναι συνεχής ενώ στα περιφερειακά δεδομένα διακριτός. Επιπρόσθετα, είναι φανερό ότι τα περισσότερα οικονομικά και δημογραφικά δεδομένα ανήκουν στην κατηγορία των περιφερειακών δεδομένων.



Διάγραμμα 1.1
Παράδειγμα γεωστατιστικών δεδομένων:
Σημεία δειγματοληψίας (αριστερά) και τιμές (δεξιά) συγκέντρωσης ψευδαργύρου στη κοιλάδα του ποταμού Μεύση στο χωριό Stein στην Ολλανδία

Στους χάρτες του Διαγράμματος 1.1 απεικονίζονται τα σημεία δειγματοληψίας και οι τιμές ψευδαργύρου που μετρήθηκαν από έναν ερευνητή στην κοιλάδα του ποταμού Μεύση κοντά στο χωριό Stein στην Ολλανδία.[1] Οι τιμές του ψευδαργύρου που έχουν καταγραφεί αποτελούν γεωστατιστικά δεδομένα καθώς θα μπορούσαν να μετρηθούν οπουδήποτε γύρω από το ποτάμι. Από το ίδιο διάγραμμα γίνεται φανερό γιατί τα γεωστατιστικά δεδομένα ονομάζονται και δεδομένα σημείου.



Διάγραμμα 1.2
Παράδειγμα περιφερειακών δεδομένων:
Διοικητική δομή της Ελλάδας (πολύγωνα) σύμφωνα με το νόμο Καλλικράτη (αριστερά) και ποσοστό μεταναστών στο τοπικό πληθυσμό κάθε δήμου (δεξιά)
Στοιχεία απογραφής 2001, πηγή Εθνική Στατιστική Υπηρεσία Ελλάδας

Στο Διάγραμμα 1.2 περιλαμβάνονται δυο χάρτες της Ελλάδας που απεικονίζουν τα διοικητικά πολύγωνα της χώρας (δήμοι) σύμφωνα με το πρόγραμμα Καλλικράτης και τα ποσοστά μεταναστών κατοίκων κάθε δήμου όπως κατεγράφησαν στην απογραφή πληθυσμού του 2001. Τα ποσοστά μεταναστών αποτελούν περιφερειακά δεδομένα και προκύπτουν αθροιστικά για κάθε πολύγωνο του χάρτη καθώς δεν είναι γνωστά τα ακριβή σημεία δειγματοληψίας των παρατηρήσεων.

[1] Τα δεδομένα περιλαμβάνονται στο πακέτο SP της R. Η παρουσίασή τους γίνεται σύμφωνα με την ανάλυση στις σελίδες 60 και 68 του βιβλίου των R. Bivand, E. Pebesma και V. Gomez-Rubio "Applied Spatial Data Analysis with R".

3) Πρότυπα σημείων (*point patterns*)

Οι δύο προηγούμενες κατηγορίες χωρικών δεδομένων έχουν κοινό χαρακτηριστικό τη σταθερή μη στοχαστική χωρική περιοχή *D*. Δηλαδή, το σύνολο *D* δεν αλλάζει σε κάθε πραγματοποίηση της διαδικασίας. Στην περίπτωση των πρότυπων σημείων η χωρική περιοχή *D* είναι τυχαία και μπορεί να αλλάζει για διαφορετικές πραγματοποιήσεις της διαδικασίας. Παράδειγμα πρότυπων σημείων θα μπορούσε να αποτελέσει η μελέτη των κεραυνών κατά τη διάρκεια μιας καταιγίδας. Η καταιγίδα δεν επηρεάζει κάθε φορά αυστηρά την ίδια περιοχή και συνεπώς δεν μπορεί να καθοριστεί επακριβώς και η περιοχή που παρατηρούνται οι πτώσεις κεραυνών. Στα πρότυπα σημείων ο ερευνητής ενδιαφέρεται να εξετάσει εάν τα συμβάντα πραγματοποιούνται με τυχαίο τρόπο ή εάν υπάρχει κάποιο χωρικό πρότυπο.

1.3 Χωρική ανάλυση, Χωρική στατιστική και Χωρική οικονομετρία

Ο ιδιαίτερος τρόπος συλλογής των χωρικών δεδομένων έχει αποτέλεσμα την εμφάνιση των χωρικών επιδράσεων (*spatial effects*) που αποτελούνται από τη χωρική εξάρτηση (*spatial dependence*) και τη χωρική ετερογένεια (*spatial heterogeneity*). Χωρική εξάρτηση σημαίνει ότι υπάρχει γεωγραφική εξάρτηση μεταξύ των παρατηρήσεων του δείγματος με συνέπεια να μην είναι ανεξάρτητες αλλά μία παρατήρηση που βρίσκεται σε κάποια συγκεκριμένη περιοχή, όπως αυτή περιγράφεται από τις γεωγραφικές της συντεταγμένες, να συσχετίζεται με κάποια άλλη παρατήρηση που βρίσκεται σε άλλη περιοχή δημιουργώντας συστάδες (*clusters*) παρατηρήσεων. Για το λόγο αυτό ονομάζεται και χωρική αυτοσυσχέτιση παραπέμποντας στον αντίστοιχο όρο της ανάλυσης χρονοσειρών με τον οποίο όμως υπάρχει σημαντική διαφορά καθώς στην ανάλυση χρονοσειρών η εξάρτηση είναι προς μόνο μία κατεύθυνση σε αντίθεση με τη χωρική εξάρτηση που μπορεί να είναι προς πολλές κατευθύνσεις. Χωρική ετερογένεια σημαίνει ότι η σχέση ανάμεσα στις εξεταζόμενες μεταβλητές δεν παραμένει σταθερή σε ολόκληρη τη γεωγραφική περιοχή αλλά μεταβάλλεται από σημείο σε σημείο.

Η παρουσία των χωρικών επιδράσεων εάν αγνοηθεί σε μια στατιστική έρευνα έχει συνέπεια τα αποτελέσματα την ανάλυσης να είναι μη αξιόπιστα. Το πρόβλημα αυτό οδήγησε στην ανάπτυξη ενός καινούριου επιστημονικού τομέα, δηλαδή της χωρικής ανάλυσης (*spatial analysis*) που διαθέτει μεθοδολογικά εργαλεία για την ανάλυση χωρικών δεδομένων. Μέσα στον ευρύτερο τομέα της χωρικής ανάλυσης έχει αναπτυχθεί η χωρική στατιστική (*spatial*

statistics) που προσπαθεί να ερμηνεύσει σχέσεις μεταξύ μεταβλητών λαμβάνοντας ύπ' όψιν τις γεωγραφικές τους ιδιότητες, δηλαδή τις θέσεις που βρίσκονται οι στατιστικές μονάδες στο χώρο. Στην πραγματικότητα επεκτείνει τις καθιερωμένες τεχνικές της στατιστικής στη μελέτη χωρικών δεδομένων αντιμετωπίζοντας τα προβλήματα που εμφανίζονται λόγω της ιδιομορφίας τους και ταυτόχρονα εκμεταλλεύεται τις γεωγραφικές τους ιδιότητες για την αποκάλυψη περεταίρω χρήσιμων πληροφοριών.

Ιστορικά, η πρώτη εφαρμογή χωρικής ανάλυσης έγινε από το γιατρό John Snow το 1854 στο Λονδίνο. Συγκεκριμένα, στο Λονδίνο της Βικτωριανής εποχής υπήρχε έξαρση θανάτων από επιδημία χολέρας. Την εποχή εκείνη ήταν άγνωστες οι αιτίες που προκαλούν την ασθένεια και οι τρόποι με τους οποίους μεταδίδεται. Η κυριαρχούσα άποψη μεταξύ των επιστημόνων της εποχής ήταν ότι η αιτία μετάδοσης της νόσου ήταν αποτέλεσμα μόλυνσης στον αέρα της περιοχής. Ο John Snow (1855) πίστευε ότι το πρόβλημα οφειλόταν στην κατανάλωση μολυσμένου νερού. Για να στηρίξει την άποψή του κατασκεύασε ένα χάρτη της περιοχή όπου υπήρχαν τα κρούσματα της νόσου στον οποίο τοποθέτησε τα σημεία που εντοπίζονταν οι θάνατοι και τις αντλίες ύδρευσης που παρείχανε νερό. Με τον τρόπο αυτό ανακάλυψε ότι στο χάρτη υπήρχαν συμπλέγματα (*clusters*) συμβάντων κοντά σε συγκεκριμένες αντλίες ύδρευσης. Η εξέταση της ποιότητας του νερού από αυτές τις αντλίες αποκάλυψε ότι ήταν μολυσμένο με αποχετευτικά λύματα και η απομόνωσή τους από το δίκτυο ύδρευσης οδήγησε στην αντιμετώπιση της επιδημίας. Ο John Snow συνέλαβε την ιδέα και τη χρησιμότητα του γεωγραφικού συστήματος πληροφοριών (*GIS*) πολλά χρόνια νωρίτερα από την ανακάλυψη των ηλεκτρονικών υπολογιστών και παρατηρώντας την ύπαρξη χωρικής αυτοσυσχέτισης στα δεδομένα του πραγματοποίησε την πρώτη χωρική στατιστική ανάλυση ώστε δικαίως να του αποδίδεται ο χαρακτηρισμός "πατέρας της χωρικής ανάλυσης".

Πέρασε περίπου ένας αιώνας από την εργασία του John Snow, με εξαίρεση κάποιες μεμονωμένες εργασίες, ώστε να αρχίσει να δίνεται έμφαση στις ιδιαιτερότητες των χωρικών δεδομένων και στην αναγκαιότητα ανάπτυξης εξειδικευμένων τεχνικών.[2] Ο Moran (1948 και 1950) με την παρουσίαση του δείκτη χωρικής αυτοσυσχέτισης *I* γνωστού με το όνομά του έδωσε την ώθηση για την ανάπτυξη της χωρικής στατιστικής. Στις αρχές του 1970 ο Βέλγος οικονομολόγος Jean Paelinck χρησιμοποίησε τον όρο χωρική οικονομετρια (*spatial econometrics*) και μίλησε για την ανάγκη δημιουργίας ενός νέου κλάδου στην οικονομετρία που να ασχολείται με τα προβλήματα που εμφανίζονται όταν εκτιμώνται οικονομετρικά

[2] Μια λεπτομερειακή ιστορική αναδρομή στην ανάπτυξη της χωρικής στατιστικής και της χωρικής οικονομετρίας μπορεί να βρεθεί στους Anselin και Griffith (1988) και Anselin (2010).

υποδείγματα με περιφερειακά δεδομένα. Στο ομώνυμο βιβλίο (Paelinck και Klaassen, 1979) παρουσίασε τα πρώτα μεθοδολογικά εργαλεία και ακολούθησαν τα βιβλία των Cliff και Ord (1981) και του Anselin (1988). Το τελευταίο θέτει ολοκληρωμένα το μεθοδολογικό πλαίσιο της χωρικής οικονομετρίας συγκεντρώνοντας μαζί εξειδικευμένους στατιστικούς ελέγχους και τεχνικές εκτίμησης υποδειγμάτων που βρίσκονταν σε διάφορες εργασίες με αποτέλεσμα να γίνει το βιβλίο σημείο αναφοράς κάθε εργασίας στη χωρική οικονομετρία.

Στο ίδιο βιβλίο ο Anselin αναφέρει τη διαφορά μεταξύ χωρικής στατιστικής και χωρικής οικονομετρίας.[3] Ειδικότερα, η χωρική οικονομετρία διακρίνεται από τη χωρική στατιστική για τους ίδιους λόγους που γίνεται η διάκριση οικονομετρίας και στατιστικής. Δηλαδή, η χωρική στατιστική οδηγείται από τα χωρικά δεδομένα (*data – driven orientation*) στα πλαίσια της διερευνητικής ανάλυσης δεδομένων (*Exploratory Data Analysis*) χωρίς να υπάρχει εκ των προτέρων κάποιο θεωρητικό υπόδειγμα που να ερμηνεύει τα αποτελέσματα σε αντίθεση με τη χωρική οικονομετρία που υπάρχει το θεωρητικό υπόδειγμα (*model – driven approach*) που πρέπει να εκτιμηθεί και να ελεγχθεί χρησιμοποιώντας τα χωρικά δεδομένα. Όμως, ενώ η χωρική στατιστική χρησιμοποιείται κυρίως για τη μελέτη φυσικών φαινομένων και η χωρική οικονομετρία για την ανάλυση οικονομικών δεδομένων οι μεθοδολογίες τους πολλές φορές αλληλοκαλύπτονται με αποτέλεσμα να μη γίνεται πάντα ορατή η διάκρισή τους.

1.4 Οι χωρικές επιδράσεις

Το βασικό χαρακτηριστικό των χωρικών δεδομένων το οποίο τα διαφοροποιεί από τα άλλα είδη στατιστικών δεδομένων που χρησιμοποιούνται σε μια οικονομετρική ανάλυση είναι η ύπαρξη των χωρικών επιδράσεων. Οι χωρικές επιδράσεις (*spatial effects*) προέρχονται από τη φύση των χωρικών δεδομένων και διακρίνονται στη χωρική εξάρτηση (*spatial dependence*) και στη χωρική ετερογένεια (*spatial heterogeneity*).

1.4.1 Χωρική εξάρτηση

Με τον όρο χωρική εξάρτηση (*spatial dependence*) εννοείται ότι υπάρχει κάποιας μορφής σχέση στη γεωγραφική περιοχή που μελετάται ένα φαινόμενο έτσι ώστε κάτι που συμβαίνει σε ένα γεωγραφικό σημείο να επηρεάζει το αποτέλεσμα σε ένα άλλο γεωγραφικό σημείο

[3] Βλέπε Anselin (1988) σελίδα 10.

(Anselin, 1988). Η έννοια της χωρικής εξάρτησης, κατά κάποιον τρόπο, εκφράζεται από το γνωστό ως πρώτο νόμο 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." Για παράδειγμα, το εισόδημα σε μια περιοχή εξαρτάται περισσότερο από το εισόδημα των γειτονικών περιοχών από ότι το εισόδημα των πιο απομακρυσμένων περιοχών. Δύο είναι οι μορφές με τις οποίες εμφανίζεται η χωρική εξάρτηση στα χωρικά δεδομένα. Η πρώτη είναι η εμφάνισή της ως ενόχληση και η δεύτερη η εμφάνισή της με ουσιαστικό τρόπο.

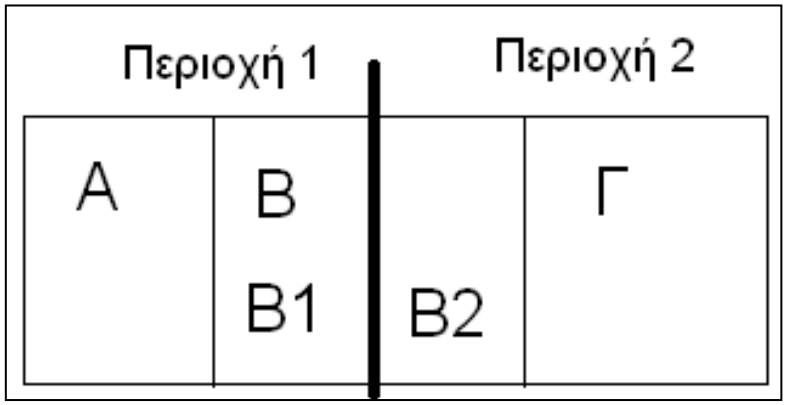
1) Χωρική εξάρτηση ως ενόχληση (*nuisance spatial dependence*)

Μια αιτία που οδηγεί σε χωρική εξάρτηση είναι η ύπαρξη σφάλματος μέτρησης στις τιμές της εξεταζόμενης μεταβλητής. Σφάλμα μέτρησης σε μια χωρική μεταβλητή μπορεί να προκύψει όταν η χωρική μονάδα στην οποία παρατηρείται ένα φαινόμενο δε συμπίπτει με τον πραγματικό χώρο στον οποίον στην πραγματικότητα γίνεται. Αυτό μπορεί να συμβεί με τα αθροιστικά δεδομένα. Παραδείγματος χάριν, όταν μελετάται η αγορά εργασίας τα διαθέσιμα στοιχεία απασχόλησης, ανεργίας και εισοδήματος μπορεί να συλλέγονται με βάση τους νομούς ή τα απογραφικά πολύγωνα αλλά η πραγματική οικονομική δραστηριότητα μίας χωρικής μονάδας να επεκτείνεται και σε γειτονικές χωρικές μονάδες. Δηλαδή, σε αυτή την περίπτωση, η χωρική εξάρτηση είναι αποτέλεσμα του τρόπου συλλογής των στατιστικών πληροφοριών, συνέπεια μη συμφωνίας των γεωγραφικών συνόρων με τα πραγματικά γεωγραφικά όρια στα οποία εξελίσσεται το φαινόμενο.

Για να γίνει κατανοητό πως μπορεί να προκύψει χωρική εξάρτηση με αθροιστικά δεδομένα υποτίθεται ότι υπάρχουν τρεις περιοχές Α, Β και Γ που αποτελούν την πραγματική χωρική κλίμακα όπου πραγματοποιούνται οι τιμές μιας μεταβλητής Υ, όπως φαίνεται στο Διάγραμμα1.3. Έστω ότι τα δεδομένα συλλέγονται από τη στατιστική υπηρεσία και αθροίζονται στα επίπεδα των περιοχών 1 και 2. Κατά συνέπεια, για τις παρατηρήσεις τις μεταβλητής Υ στις δύο περιοχές άθροισης 1 και 2 θα ισχύουν $Y_1 = Y_A + Y_{B1}$ και $Y_2 = Y_\Gamma + Y_{B2}$. Εφόσον η συλλογή των δεδομένων εμπεριέχει σφάλματα μέτρησης αυτά θα αντανακλώνται και στις τιμές των παρατηρήσεων Y_1 και Y_2 που θα παρουσιάζουν χωρική εξάρτηση. Τα σφάλματα μέτρησης αυτού του είδους στις τιμές των μεταβλητών θα οδηγήσουν και σε χωρική εξάρτηση στα σφάλματα ενός γραμμικού υποδείγματος παλινδρόμησης εφόσον οι μεταβλητές αυτές χρησιμοποιηθούν στην εκτίμησή του με

συνέπεια την υποεκτίμηση των διακυμάνσεων των εκτιμητών των συντελεστών του υποδείγματος. Επιπρόσθετα, μια ακόμη αιτία που μπορεί να προκαλέσει την εμφάνιση χωρικής εξάρτησης στα σφάλματα ενός γραμμικού υποδείγματος παλινδρόμησης είναι η παράλειψη κάποιας ανεξάρτητης ή μη μετρήσιμης μεταβλητής που ερμηνεύει την εξαρτημένη μεταβλητή όταν αυτή δεν είναι χωρικά ανεξάρτητη λόγω σφαλμάτων μέτρησης.

Αυτή η μορφή χωρικής εξάρτησης συχνά αποκαλείται ενοχλητική (*nuisance*) και συνήθως εμφανίζεται στα σφάλματα του γραμμικού υποδείγματος παλινδρόμησης. Στην περίπτωση αυτή το πρόβλημα δεν θεωρείται ουσιαστικό για την εξειδίκευση του υποδείγματος και την ερμηνεία της εξαρτημένης μεταβλητής και ο ερευνητής που «ενοχλείται» επιδιώκει να το αντιμετωπίσει με τη βοήθεια καταλλήλων οικονομετρικών τεχνικών οι οποίες θα το εξαλείψουν και θα προσδιορίσουν τις σωστές διακυμάνσεις των εκτιμητών όπως συμβαίνει στην ανάλυση χρονοσειρών.



Διάγραμμα 1.3
Χωρική εξάρτηση αποτέλεσμα άθροισης (Παράδειγμα από Anselin 1988)

2) Ουσιαστική χωρική εξάρτηση (*Substantive spatial dependence*)

Μια δεύτερη αιτία που προκαλεί χωρική εξάρτηση οφείλεται στην αλληλεπίδραση (*spatial interaction*) μεταξύ των χωρικών μονάδων. Η τοποθεσία και η απόσταση είναι σημαντικές και οδηγούν σε αλληλεξαρτήσεις στο χώρο. Αυτό μπορεί να συμβεί όταν, παραδείγματος χάριν, υπάρχει ανταγωνισμός στις επιχειρηματικές δραστηριότητες ανάμεσα σε γειτονικές χωρικές μονάδες έτσι ώστε όταν ένας επιχειρηματίας προβαίνει στη μείωση της τιμής ενός προϊόντος σε μια περιοχή οι επιχειρηματίες στις γειτονικές περιοχές προβαίνουν στην ίδια ενέργεια με αποτέλεσμα κάτι που παρατηρείται σε ένα γεωγραφικό σημείο να οφείλεται σε κάτι που έχει γίνει σε κάποιο άλλο γεωγραφικό σημείο. Με αυτή τη διαδικασία προκύπτει ένα αποτέλεσμα διάχυσης (*spillover effect*) που επηρεάζει όλες τις χωρικές μονάδες. Η αιτία αυτή

χωρικής εξάρτησης είναι πιο ουσιαστική διότι υποδηλώνει την ύπαρξη κάποιας χωρικής διαδικασίας (*spatial process*) που οδηγεί στην εξάρτηση η οποία πρέπει να ληφθεί υπ' όψιν από τον ερευνητή όταν εξειδικεύει ένα οικονομετρικό υπόδειγμα για την ερμηνεία του φαινομένου.

Η πλήρη μελέτη του χαρακτηριστικού της χωρικής εξάρτησης απαιτεί τη γνώση της από κοινού συνάρτησης κατανομής της εξεταζόμενης μεταβλητής για όλα τα γεωγραφικά σημεία στα οποία παρατηρείται. Επειδή είναι δύσκολη η γνώση της από κοινού συνάρτησης κατανομής συνήθως η χωρική εξάρτηση μελετάται στην ασθενέστερή της μορφή που είναι η χωρική αυτοσυσχέτιση για την οποία χρειάζονται μόνο οι ροπές της κατανομής. Κατά συνέπεια, μια τυχαία μεταβλητή *y* θα παρουσιάζει χωρική αυτοσυσχέτιση όταν ικανοποιείται η συνθήκη:

$$Cov\left(y_i, y_j\right) = E\left(y_i, y_j\right) - E\left(y_i\right)E\left(y_j\right) \neq 0, \quad \text{για } i \neq j$$

όπου *i* και *j* είναι τα σημεία δύο περιοχών στον εξεταζόμενο γεωγραφικό χώρο και y_i και y_j οι τιμές της μεταβλητής σε αυτές τις περιοχές. Αποτέλεσμα της ύπαρξης χωρικής αυτοσυσχέτισης είναι η δημιουργία συστάδων (*clusters*) των τιμών των παρατηρήσεων και όχι η τυχαία κατανομή τους επάνω σε ένα γεωγραφικό χάρτη. Ειδικότερα, όταν οι υψηλές τιμές της μεταβλητής εντοπίζονται κοντά σε υψηλές τιμές και οι χαμηλές τιμές γειτονεύουν με χαμηλές τιμές σχηματίζοντας συστάδες υψηλών και χαμηλών τιμών, τότε υπάρχει θετική χωρική αυτοσυσχέτιση. Αντίθετα, όταν οι υψηλές τιμές της μεταβλητής βρίσκονται κοντά σε χαμηλές και οι χαμηλές τιμές κοντά σε υψηλές τότε υπάρχει αρνητική χωρική αυτοσυσχέτιση. Υπάρχουν ομοιότητες με τη χρονική αυτοσυσχέτιση που εμφανίζεται στην ανάλυση χρονοσειρών όμως στις χρονοσειρές η σχέση εξάρτησης είναι μόνο προς μία κατεύθυνση (η προηγούμενη χρονικά τιμή της χρονοσειράς επηρεάζει την επόμενη) ενώ στο χώρο προς πολλές κατευθύνσεις.

1.4.2 Χωρική ετερογένεια

Η δεύτερη μορφή χωρικών επιδράσεων είναι η χωρική ετερογένεια (*spatial heterogeneity*). Ο όρος χωρική ετερογένεια αναφέρεται στην έλλειψη σταθερότητας στη σχέση μεταξύ των μεταβλητών στις διαφορετικές γεωγραφικές περιοχές. Το φαινόμενο αυτό είναι πολύ συνηθισμένο όταν μελετώνται οικονομικές και δημογραφικές μεταβλητές και οφείλεται στις πληθυσμιακές και εισοδηματικές ανισότητες μεταξύ των περιοχών, στη διαφορετική έκταση

που έχουν οι απογραφικές περιοχές και γενικότερα στην ανομοιογένεια των μεταβλητών στο χώρο. Όπως και με τη χωρική εξάρτηση έτσι και η χωρική ετερογένεια μπορεί να εμφανιστεί με τη μορφή ενόχλησης η με ουσιαστικό τρόπο.

1) Χωρική ετερογένειας ως ενόχληση (*nuisance spatial heterogeneity*)

Στην πιο απλή περίπτωση, η χωρική ετερογένεια παρουσιάζεται μέσω του προβλήματος της ετεροσκεδαστικότητας στα εκτιμηθέντα σφάλματα ενός υποδείγματος παλινδρόμησης. Δηλαδή, τα σφάλματα έχουν διαφορετική συμπεριφορά από περιοχή σε περιοχή. Η ετεροσκεδαστικότητα (έλλειψη σταθερής διακύμανσης) στα σφάλματα του γραμμικού υποδείγματος παλινδρόμησης είναι γνωστό, όπως άλλωστε συμβαίνει και με την απουσία ανεξαρτησίας, ότι οδηγεί στον προσδιορισμό λανθασμένων διακυμάνσεων των εκτιμητών και λανθασμένη στατιστική συμπερασματολογία. Αιτίες εμφάνισης αυτής της μορφής χωρικής ετερογένειας μπορούν να θεωρηθούν και πάλι τα σφάλματα μέτρησης στις τιμές των μεταβλητών, η παράλειψη ανεξάρτητων μεταβλητών και η λανθασμένη εξειδίκευση του υποδείγματος. Όταν εντοπίζεται ενοχλητική χωρική ετερογένεια ο ερευνητής προσπαθεί να την αντιμετωπίσει με χρήση οικονομετρικών τεχνικών και μεθόδων εκτίμησης.

2) Ουσιαστική χωρική ετερογένεια (*substantive spatial heterogeneity*)

Η μορφή της ουσιαστικής χωρικής ετερογένειας είναι πιο σύνθετη. Συγκεκριμένα, η ένταση αλλά και η συναρτησιακή εξάρτηση της σχέσης μεταξύ δύο ή περισσότερων μεταβλητών μπορεί να μην είναι ίδιες σε όλες τις περιοχές. Στα πλαίσια ενός υποδείγματος παλινδρόμησης η ουσιαστική χωρική ετερογένεια θα μπορούσε να παρουσιαστεί με τη σχέση:

$$y_i = f_i\left(x_{1i}, x_{2i}, \ldots, x_{ki}, \beta_{1i}, \beta_{2i}, \ldots \beta_{ki}, \varepsilon_i\right)$$

όπου y είναι η εξαρτημένη μεταβλητή, $x_1, x_2, \ldots, x_k$ οι ανεξάρτητες μεταβλητές, $\beta_1, \beta_2, \ldots \beta_k$ οι συντελεστές του υποδείγματος και ε_i τα σφάλματα στην i περιοχή. Η ουσιαστική χωρική ετερογένεια θα μπορούσε να χωριστεί σε δύο τύπους, δηλαδή στη διακριτή χωρική ετερογένεια (*discrete spatial heterogeneity*) και στη συνεχή χωρική ετερογένεια (*continuous spatial heterogeneity*).

Διακριτή χωρική ετερογένεια σημαίνει ότι οι παράμετροι του υποδείγματος σε κάποιες συγκεκριμένες γεωγραφικές περιοχές έχουν διαφορετικές τιμές από ότι στις άλλες περιοχές. Παραδείγματος χάριν, οι τιμές των παραμέτρων μπορεί να διαφοροποιούνται ανάλογα με το

εάν μια περιοχή είναι αστική ή αγροτική ή εάν ανήκει στο βόριο ή στο νότιο μέρος μιας χώρας. Στις περιπτώσεις αυτές το πρόβλημα αντιμετωπίζεται με την εισαγωγή στο υπόδειγμα ψευδομεταβλητών (*dummy variables*) που καθορίζουν τη γεωγραφική θέση κάθε παρατήρησης και που οδηγούν στην εκτίμηση διαφορετικών σταθερών όρων ή και συντελεστών των ανεξάρτητων μεταβλητών.

Η συνεχής χωρική ετερογένεια είναι πιο σύνθετη στο να αντιμετωπιστεί και να ερμηνευθεί και η χωρική οικονομετρία προτείνει ειδική μεθοδολογία. Ο τύπος αυτός ετερογένειας συνεπάγεται ότι οι συντελεστές του υποδείγματος διαφοροποιούνται σε όλες τις περιοχές και προσδιορίζονται για κάθε γεωγραφικό σημείο ως αποτέλεσμα εφαρμογής κάποιας συνάρτησης των γεωγραφικών συντεταγμένων.

1.5 Σχόλια

1) Δεν είναι υποχρεωτικό να εμφανίζονται ταυτόχρονα και οι δύο μορφές των χωρικών επιδράσεων σε μια ανάλυση.
2) Λόγω του γεγονότος ότι τα χωρικά δεδομένα είναι διαστρωματικά δεδομένα πολλές φορές δεν γίνεται ξεκάθαρο ποια μορφή χωρικών επιδράσεων εμφανίζεται.
3) Από τις δύο μορφές χωρικών επιδράσεων η χωρική εξάρτηση είναι η πιο δύσκολη να αντιμετωπιστεί και εάν αγνοηθεί μπορεί να έχει δραματικές συνέπειες στην αξιοπιστία των αποτελεσμάτων μίας οικονομετρικής ανάλυσης. Στην κατεύθυνση αντιμετώπισης των προβλημάτων που εμφανίζονται από τη χωρική εξάρτηση συμβάλλουν αποφασιστικά οι εξειδικευμένες τεχνικές της χωρικής οικονομετρίας.
4) Η δεύτερη μορφή χωρικών επιδράσεων, δηλαδή η χωρική ετερογένεια, είναι πιο εύκολο να αντιμετωπιστεί στα πλαίσια των μεθόδων της κλασσικής οικονομετρίας, ιδιαίτερα όταν εμφανίζεται το πρόβλημα της ετεροσκεδαστικότητας. Εντούτοις, και σε αυτή την περίπτωση η χωρική οικονομετρία μπορεί να βοηθήσει στο να προκύψουν πιο αξιόπιστα αποτελέσματα και περισσότερες πληροφορίες στον ερευνητή.