



Πανεπιστήμιο Αιγαίου

Χωρική Ανάλυση

Ενότητα 1α: Χωρικά δεδομένα και στόχοι χωρικής στατιστικής ανάλυσης

Κυριακίδης Φαίδων

Τμήμα Γεωγραφίας

Άδειες Χρήσης

Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.

Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.

Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.

Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Γεωχωρικά Δεδομένα και Στατιστική Ανάλυση

Φαίδων Κυριακίδης

Καθηγητής

phkyriakidis@geo.aegean.gr



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
Λόφος Πανεπιστημίου, 81100 Μυτιλήνη

Χωρική Ανάλυση

ΤΜΗΜΑ ΓΕΩΓΡΑΦΙΑΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΓΕΩΓΡΑΦΙΑ ΚΑΙ ΕΦΑΡΜΟΣΜΕΝΗ ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ

Εισαγωγή

(Γεω)Χωρικά Δεδομένα



Ορισμός

Γεω-αναφερόμενα δεδομένα ή γεω-δεδομένα (georeferenced measurements):
κάθε δεδομένο ή μέτρηση σχετίζεται με μια θέση, σημείο, περιοχή ή οντότητα
(αντικείμενο) στο γεωγραφικό, ή σε άλλο, χώρο

- ▶ συνεχής ή διακριτή κλίμακα μέτρησης ενός χαρακτηριστικού (attribute) σε **μια** θέση
π.χ. χημική περιεκτικότητα, εδαφικός τύπος, εκδήλωση νόσου
- ▶ θέσεις μετρήσεων που κατανέμονται κανονικά ή όχι στο χώρο, π.χ. μετρήσεις
δισπαρμένες (scattered) στο χώρο ή κατανεμημένες σε κανονική κίνναβο (regular
lattice): ο χώρος που αντιστοιχεί σε κάθε μέτρηση ονομάζεται *υποστήριγμα*
(support) π.χ. σημείο, pixel, πολύγωνο, voxel
- ▶ συνήθως, υπάρχει και μια χρονική διάσταση στις γεωγραφικές μετρήσεις
μετρήσεις δυναμικών διεργασιών, χωρο-χρονικό υποστήριγμα

Στόχοι του μαθήματος αυτού

- ▶ σύντομη επισκόπηση των διαφόρων τύπων χωρικών δεδομένων
- ▶ παρουσίαση του ρόλου της χωρικής στατιστικής στη χωρική ανάλυση



Τί Είναι η Χωρική Ανάλυση;

Εναλλακτικοί ορισμοί

- ▶ Η επεξεργασία χωρικών δεδομένων, συνήθως σε ένα Γεωγραφικό Σύστημα Πληροφοριών (ΓΣΠ), που συχνά περιλαμβάνει την απεικόνιση/οπτικοποίηση, περιγραφή και ανίχνευση
- ▶ Η χρησιμοποίηση της στατιστικής ανάλυσης για να διερευνήσει κανείς αν οι τιμές μιας χωρικής μεταβλητής είναι 'τυπικές' ή 'αναμενόμενες' με βάση κάποιο στατιστικό μοντέλο
- ▶ Η κατασκευή και χρησιμοποίηση χωρικών μοντέλων, ή καλύτερα μοντέλων των οποίων τα εισερχόμενα (inputs) είναι χωρικά δεδομένα. Στην ανθρωπογεωγραφία, π.χ. τέτοια μοντέλα χρησιμοποιούνται για την πρόβλεψη ροής πληθυσμού, ενώ στη φυσική γεωγραφία αντίστοιχα μοντέλα χρησιμοποιούνται για τη δυναμική προσομοίωση φυσικών διεργασιών, όπως η βροχόπτωση

Στάδια στη Χωρική Ανάλυση



Διερευνητική Ανάλυση

- ▶ οπτικοποίηση ή/και περίληψη δεδομένων με περιγραφικά στατιστικά
- ▶ διερεύνηση τυχών υπο-πληθυσμών (sub-populations), ασυνήθιστων τιμών (outliers), ροπών (trends), συσχέτισης με παρακείμενες τιμές τις ίδιας ή άλλης χωρικής μεταβλητής

Επιβεβαιωτική ή Επαγωγική Ανάλυση

- ▶ κατασκευή παραμετρικών ή μη παραμετρικών μοντέλων χωρικών διεργασιών που να εξηγούν τη χωρική κατανομή ενός χαρακτηριστικού ή μιας μεταβλητής
- ▶ εκτίμηση παραμέτρων από τα δεδομένα και πιθανή πρόβλεψη τιμών της ίδιας ή άλλης μεταβλητής

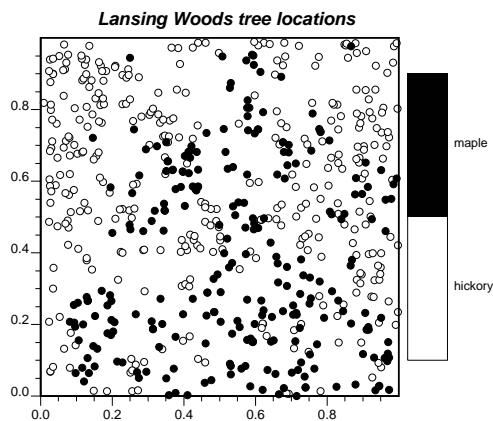
Σημειώσεις

- ▶ τα όρια των παραπάνω σταδίων δεν είναι πάντοτε σαφή ...
- ▶ οι διάφορες επεξεργασίες, π.χ. φιλτράρισμα, χωρική παρεμβολή, στις οποίες υπόκεινται τα χωρικά δεδομένα επηρεάζουν τα όποια συμπεράσματα εξάγονται από την ανάλυσή τους

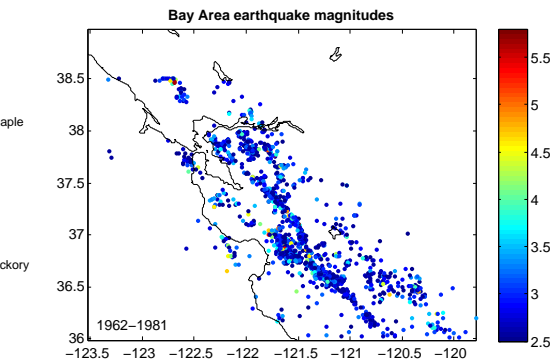


Χωρικά Σημειακά Πρότυπα -- Spatial Point Patterns

- ▶ σύνολο σημειακών θέσεων με καταγεγραμμένα συμβάντα, π.χ. θέσεις δένδρων, εκδήλωσης νόσου ή εγκλήματος
- ▶ οι θέσεις είναι δυνατόν να καταγράφουν όλα τα πιθανά συμβάντα – χαρτογραφημένο (*mapped*) σημειακό πρότυπο, ή ένα υποσύνολο αυτών – δειγματοληπτικό (*sampled*) σημειακό πρότυπο
- ▶ όταν σε κάθε θέση έχει μετρηθεί και κάποιο χαρακτηριστικό, π.χ. διάμετρος κορμού δέντρου, μέγεθος σεισμού, τότε αναφερόμαστε σε ένα σημαδεμένο σημειακό πρότυπο (*marked point pattern*)



Φ. Κυριακίδης (Παν. Αιγαίου)



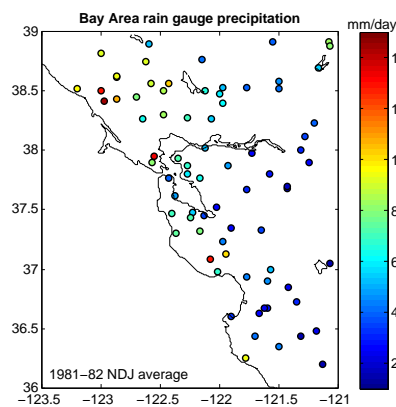
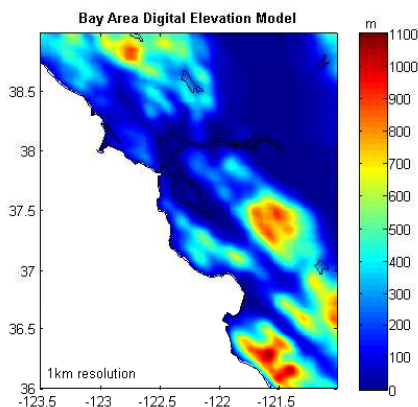
Χωρική Ανάλυση

Γεωχωρικά Δεδομένα & Στατιστική Ανάλυση

Μεταβλητές Συνεχούς Χωρικής Κατανομής – Πεδία



- ▶ μεταβλητές που κατανέμονται συνεχώς στο χώρο (*continuous spatial variables*), π.χ. θερμοκρασία επιφανείας, τοπογραφικό υψόμετρο, πληθυσμιακή πυκνότητα
- ▶ κλίμακα ή επίπεδο μέτρησης: αναλογικό, π.χ. θερμοκρασία, ή ονομαστικό, π.χ. κατηγορίες εδαφών ή χρήσεων Γης
- ▶ συχνά, οι μετρήσεις των μεταβλητών αυτών γίνονται σε λίγα και χωρικά διεσπαρμένα σημεία

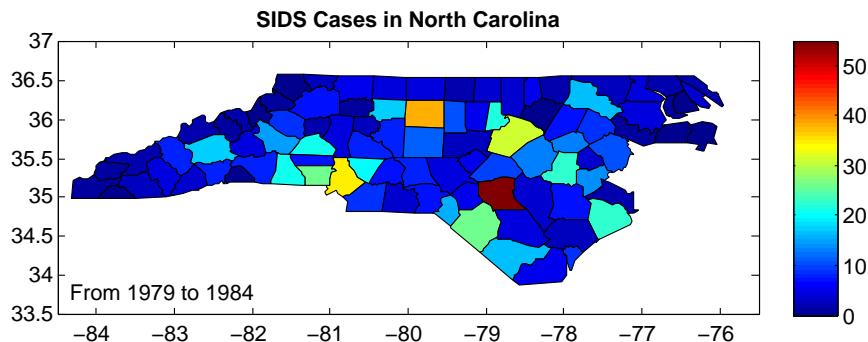


Υποσημείωση: Τα σημεία μέτρησης υψόμετρου που απαρτίζουν ένα Ψηφιακό Μοντέλο Αναγλύφου δεν είναι ακριβώς διατεταγμένα σε μια κανονική κάρναβο. Ακόμη κι αν αυτό ήταν αληθές, το ουσιαστικό είναι ότι το υψόμετρο μπορεί να μετρηθεί σε οποιαδήποτε σημείο της Γης



Μεταβλητές Διακριτής Χωρικής Κατανομής – Πολύγωνα

- ▶ μεταβλητές που λαμβάνουν τιμές σε ένα πεπερασμένο αριθμό επιφανειών ή ζωνών (areas or zones), π.χ. διοικητικές περιοχές, pixels δορυφορικών εικόνων
- ▶ κλίμακα ή επίπεδο μέτρησης: αναλογικό, π.χ. μέση τιμή ρύπων, μέσο εισόδημα, ή ονομαστικό, π.χ. δεσπόζουσα κατηγορία εδάφους, κοινωνικο-οικονομικό στρώμα
- ▶ συνήθως, όλες οι δυνατές μετρήσεις έχουν ληφθεί και συνεπώς δεν υφίσταται η έννοια τιμών της μεταβλητής μεταξύ των ορίων των υφιστάμενων περιοχών (εκτός αν υπάρχουν ελλειπίες μετρήσεις)



Υποσημείωση: Ο διαχωρισμός μεταξύ των μεταβλητών συνεχούς και διακριτής χωρικής κατανομής δεν είναι πάντα σαφής, ειδικά όταν οι τελευταίες ορίζονται ως χωρικό συνάθροισμα των προηγούμενων

Δεδομένα Χωρικής Αλληλεπίδρασης – Spatial Interaction Data



Γνωστά και ως δεδομένα δικτύων (network data)

- ▶ δεδομένα μεταβλητών που αναφέρονται σε ζεύγη σημείων ή περιοχών, δηλαδή 'ροές' από αφετηρίες (origins) σε προορισμούς (destinations), π.χ. αριθμός ασθενών από οικίες σε νοσοκομεία
- ▶ λιγότερο απτές ροές μπορούν να ορισθούν, όπως η ροή πληροφοριών μεταξύ κόμβων στο διαδίκτυο

Στόχοι χωρικής ανάλυσης

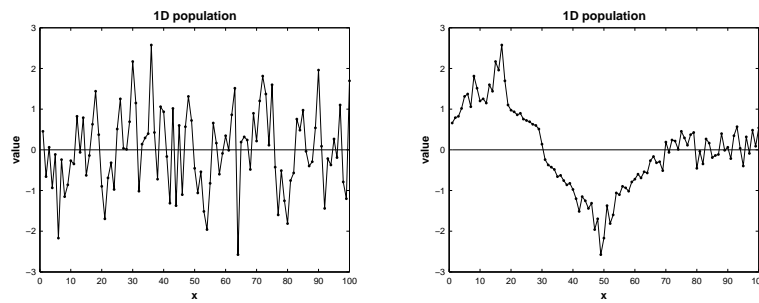
- ▶ μοντελοποίηση των προτύπων ροής = επεξήγηση των παρατηρούμενων ροών ως συνάρτηση άλλων μεταβλητών, π.χ. αριθμός ταξιδιών σε σχέση με το εισόδημα
- ▶ προβλήματα γεωγραφικής χωροθέτησης (location-allocation) και γενικότερα χωρικής βελτιστοποίησης (spatial optimization), συνήθως εμπεριέχουν τέτοιου τύπου δεδομένα

Υποσημείωση: Η ανάλυση αυτού του τύπου δεδομένων δεν καλύπτεται σε αυτό το μάθημα



Γιατί Χωρική Στατιστική;

Δύο μονοδιάστατες κατανομές τιμών μιας μεταβλητής με το ίδιο ακριβώς ιστογράμμα:



Ανεπάρκεια κλασικής στατιστικής

Η μονομεταβλητή στατιστική ανάλυση μέσω του μέσου όρου, της διακύμανσης ή και του ιστογράμματος, δεν επαρκεί για να ποσοτικοποιήσει και να περιγράψει χωρικά πρότυπα (spatial patterns). Η χωρική διάταξη των τιμών της μεταβλητής παίζει σημαντικότατο ρόλο

Χωρική αυτοσυσχέτιση μια συνιστώσα του χωρικού προτύπου

Οι τιμές της μεταβλητής που μετρήθηκαν σε 'παρακείμενα' υποστηρίγματα είναι συνήθως πιο 'παρόμοια' από τιμές που μετρήθηκαν σε 'απομακρυσμένα' υποστηρίγματα. Αυτός θεωρείται και ο 1ος νόμος της Γεωγραφίας ή νόμος του Tobler

Ο Ρόλος της Χωρικής Στατιστικής στη Χωρική Ανάλυση



Για χωρικά σημειακά πρότυπα – spatial point patterns

- ▶ ανίχνευση συγκέντρωσης (clustering) ή κανονικότητας (regularity), σε αντίθεση με την πλήρη έλλειψη προτύπου, στις θέσεις συμβάντων στο χώρο και το χρόνο
- ▶ σε περίπτωση ανίχνευσης συγκέντρωσης, τότε το επόμενο βήμα είναι η διερεύνηση πιθανών σχέσεων μεταξύ των συγκεντρώσεων αυτών και παρακείμενων 'πηγών' (sources) ή άλλων σχετικών μεταβλητών

Για μεταβλητές συνεχούς χωρικής κατανομής

- ▶ μοντελοποίηση της χωρικής κατανομής της μεταβλητής από σημειακά δεδομένα
- ▶ πρόβλεψη τιμών μιας μεταβλητής σε θέσεις χωρίς μετρήσεις, πιθανώς λαμβάνοντας υπ όψη και δεδομένα άλλων σχετικών μεταβλητών

Για δεδομένα επιφάνειας – areal data

- ▶ διερεύνηση χωρικών προτύπων και τάσεων (trends) ή σχέσεων με άλλες μεταβλητές, π.χ. συσχέτιση ποσοστών νόσου με κοινωνικο-οικονομικά δεδομένα
- ▶ δεν νοείται η πρόβλεψη νέων τιμών της μεταβλητής σε πολύγωνα, εκτός αν απαιτείται η προσθήκη ελλειπών τιμών ή η διόρθωσή τους



Χωρική ή Μη Χωρική Στατιστική;

Κλασική στατιστική

- ▶ τα δεδομένα προέρχονται από ανεξάρτητες τυχαίες μεταβλητές (independent random variables) με την ίδια στατιστική κατανομή (statistical distribution)
- ▶ οι στατιστικοί έλεγχοι υποθέσεων (statistical hypothesis tests) βασίζονται στην παραπάνω παραδοχή, γεγονός που προκαλεί προβλήματα στην ανάλυση χωρικών δεδομένων

Χωρική στατιστική

- ▶ πολυμεταβλητή στατιστική σε ένα χωρικό ή χωροχρονικό πλαίσιο: τα δεδομένα προέρχονται από χωρικά ή χωροχρονικά εξαρτημένες μεταβλητές με πιθανώς διαφορετική στατιστική κατανομή – αυτό σημαίνει ότι κάθε δεδομένο δεν είναι μια ανεξάρτητη πηγή πληροφορίας
- ▶ η χωρική συσχέτιση μοντελοποιείται και λαμβάνεται υπ' όψην στον προσδιορισμό διαστημάτων εμπιστοσύνης και στους στατιστικούς ελέγχους

Τα παραπάνω δεν σημαίνουν ότι η κλασική στατιστική δεν έχει θέση στη χωρική ανάλυση.
Όμως πρέπει να θυμάται κανείς ότι υπάρχουν προβλήματα στους σχετικούς στατιστικούς ελέγχους . . .

Λογισμικά Πακέτα για Χωρική Στατιστική Ανάλυση



Σε περιβάλλον ΓΣΠ

- ▶ ESRI's Spatial Analyst, Geostatistical Analyst . . .
- ▶ όταν κάποια συγκεκριμένη στατιστική ανάλυση απουσιάζει από το ΓΣΠ, τότε είναι δυνατή η "close" ή "loose" σύζευξη με εξειδικευμένα στατιστικά πακέτα

Ολοκληρωμένα στατιστικά πακέτα

Το R και το Matlab είναι τα πλέον διαδεδομένα στη Γεωγραφία – το πρώτο είναι ανοικτού κώδικα, το δεύτερο είναι ακριβό

Άλλα πακέτα

- ▶ το IDL και το Matlab είναι τα πλέον διαδεδομένα στην Τηλεπισκόπηση και στην επεξεργασία σήματος/εικόνας γενικότερα
- ▶ χωρικά σημειακά πρότυπα: CrimeStat, δεδομένα επιφανειών: GeoDa και Spatial Analysis in Macroecology (SAM), σημειακά δεδομένα: Geostatistical Software Library (GSLIB) και Stanford Geostatistical Modeling Software (SGEMS)

Η πρόσβαση σε κώδικα που είναι γραμμένος σε μια απλή και κατανοητή γλώσσα προγραμματισμού είναι καίρια για την ανάπτυξη νέας έρευνας και τεχνολογίας. . .



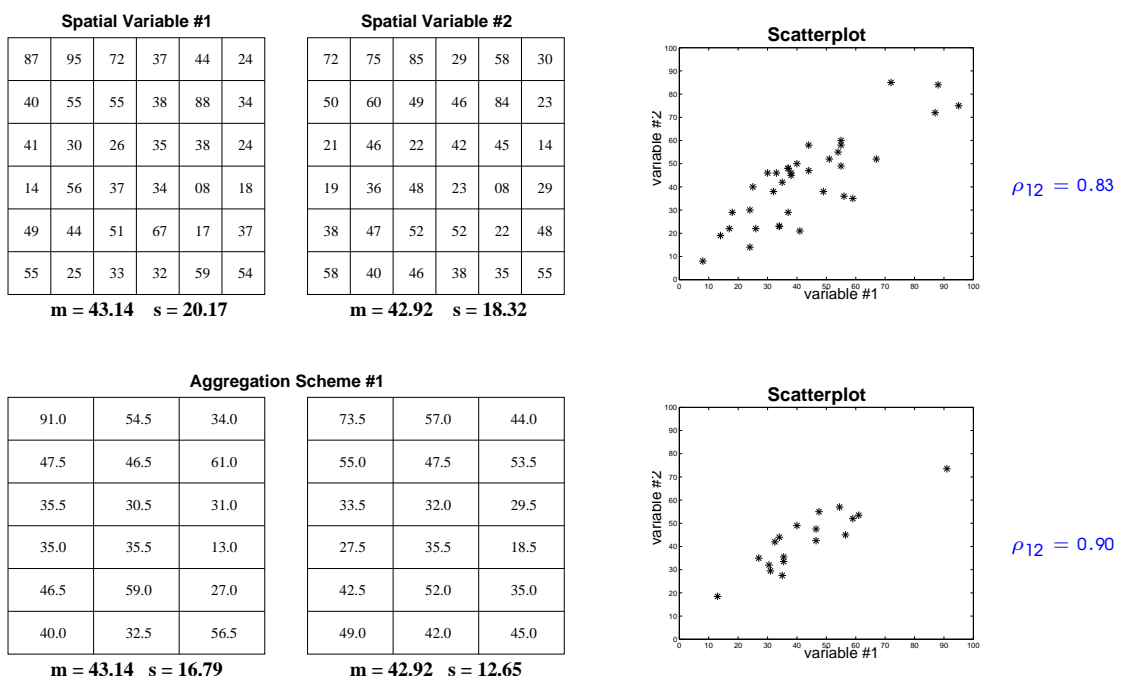
Γενικές Παρατηρήσεις Για Τα Χωρικά Δεδομένα

- ▶ διαφορές από δεδομένα χρονοσειρών: (1) μη κανονική δειγματοληψία, (2) έλλειψη δεικτοδότησης (indexing), χωρίς τη δυνατότητα ορισμού παρελθόντος, παρόντος και μέλλοντος – (3) αυτο- και ετερο-συσχέτιση μεταξύ μεταβλητών, της οποίας ο βαθμός μπορεί να μεταβάλλεται ανάλογα με τη διεύθυνση (ανισοτροπία)
- ▶ δεδομένα από *πολλαπλές πηγές* με διαφορετική χωρική (ή και χρονική) διακριτική ικανότητα (resolution)
- ▶ δεδομένα που συχνά αφορούν *αθροίσματα* σημειακών (ατομικών) τιμών σε αυθαίρετα ορισμένες περιοχές. Τα χωρικά στατιστικά αθροιστικών δεδομένων, όμως, *διαφέρουν* από τα χωρικά στατιστικά των ατομικών (επιμέρους) τιμών
 1. πρόβλημα μεταβαλλόμενης χωρικής μονάδας – Modifiable Area Unit Problem (MAUP)
 2. πρόβλημα οικολογικής συμπερασματολογίας – Ecological Inference Problem (EIP)
- ▶ *επιδράσεις ορίων*: μετρήσεις που λαμβάνονται κοντά στα όρια της περιχής μελέτης μπορεί να επηρεάζονται από διαφορετικές χωρικές διεργασίες απ' ότι μετρήσεις στο εσωτερικό της περιοχής
- ▶ στα μοντέλα χωρικών διεργασιών γίνεται συχνά ο διαχωρισμός μεταξύ επιδράσεων *πρώτου* και *δεύτερου* βαθμού, δηλ., μεταξύ περιβαλλοντικών επιδράσεων και επιδράσεων αλληλεπίδρασης.
Ένας τέτοιος διαχωρισμός δεν είναι πάντα εφικτός με βάση μόνο τις μετρήσεις

Modifiable Area-Unit Problem: Aggregation Effect



Two spatial variables and their univariate/bivariate statistics

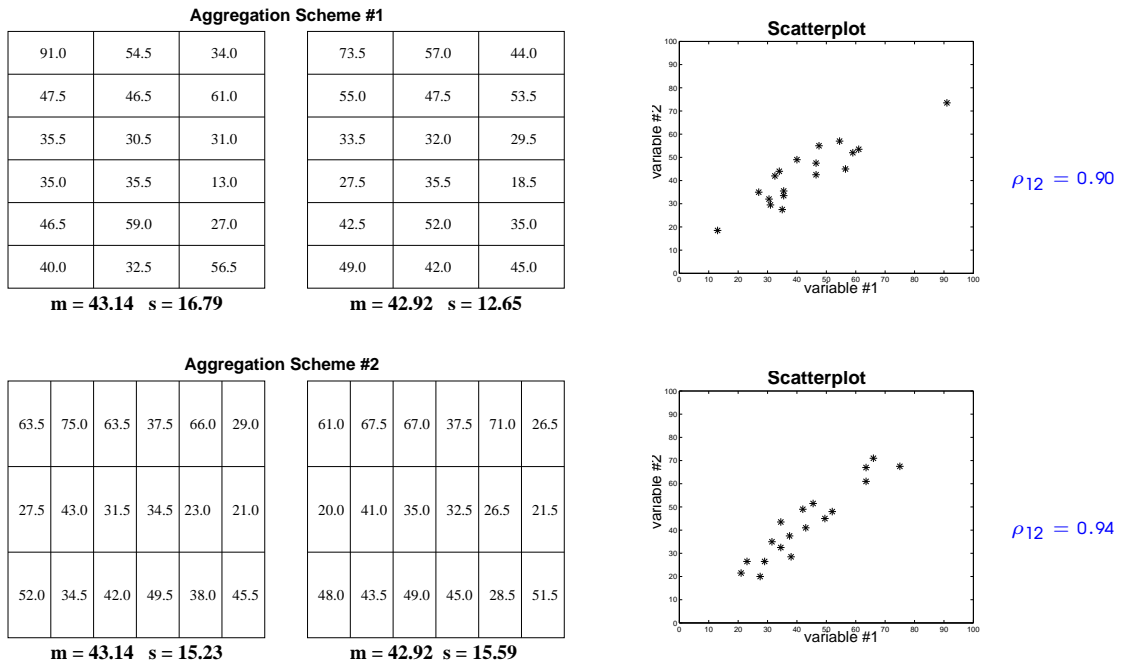


Statistics and relationships between spatial attributes depend on *aggregation extent*



Modifiable Area-Unit Problem: Zonation Effect

Upscaling spatial variables using two different aggregation schemes

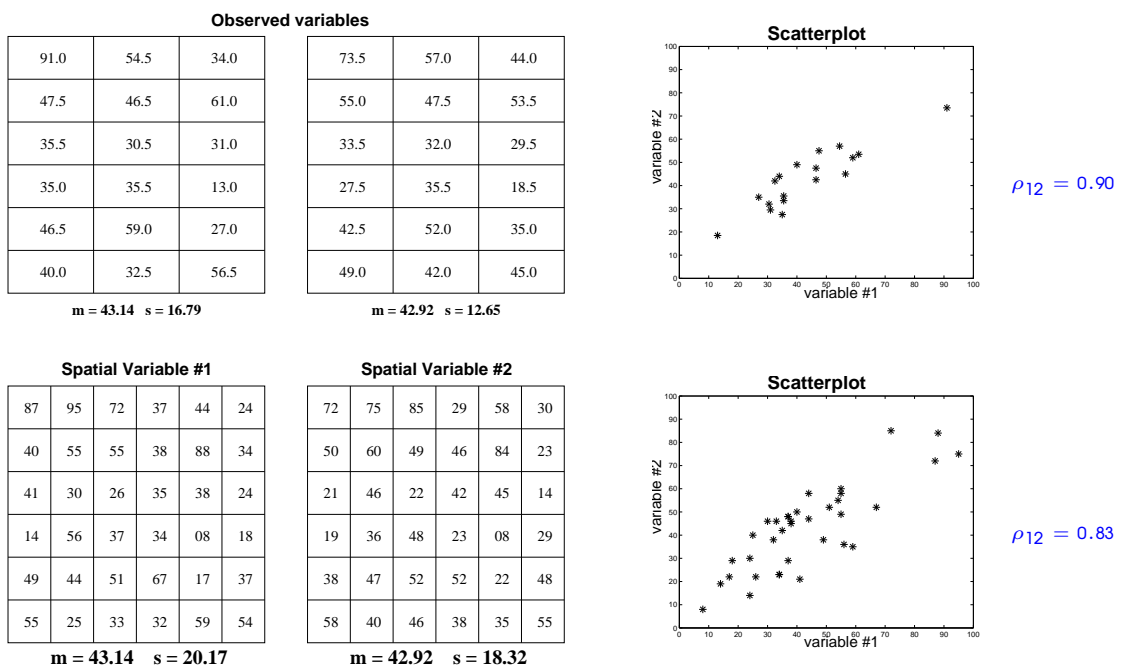


For a given aggregation extent, statistics and relationships between spatial attributes depend on which individual values are aggregated and how

Ecological Inference Problem I



Downscaling spatial variables

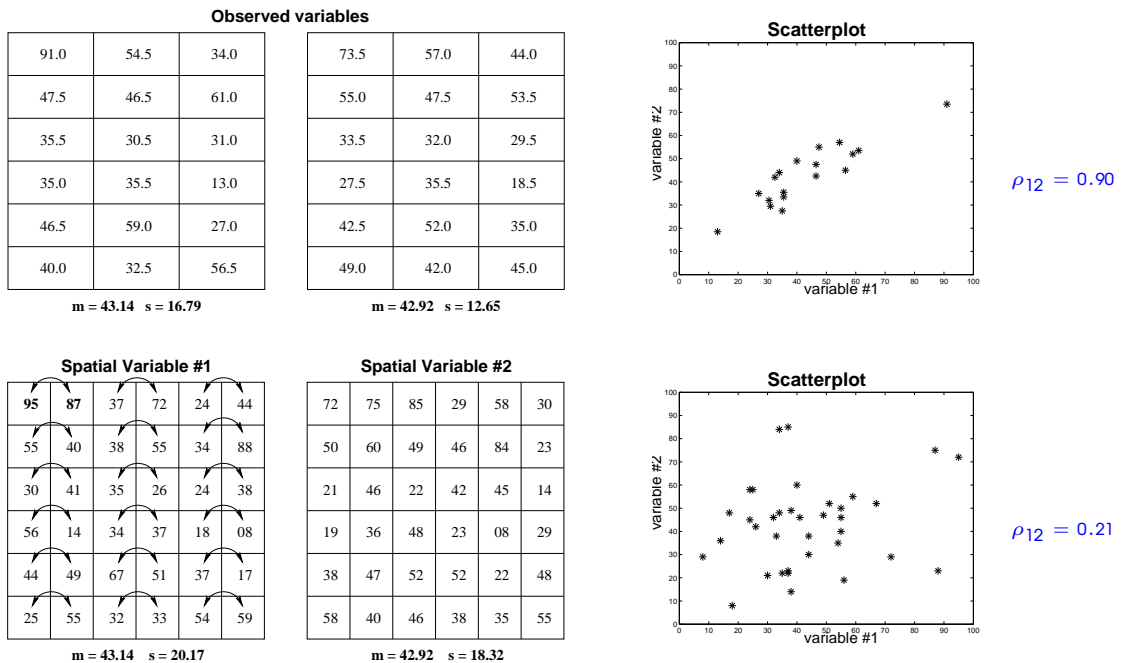


Statistics and relationships between spatial variables at a finer spatial resolution are *different* than those derived at the original coarse resolution



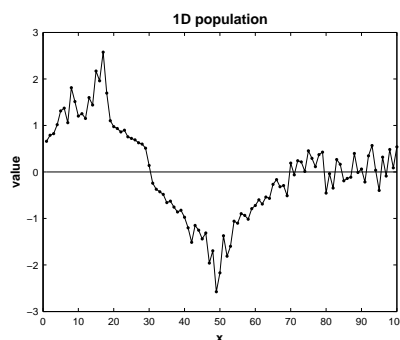
Ecological Inference Problem II

Under-determined inverse problem



Multiple combinations of fine spatial resolution attribute values can lead to the same aggregate values at a coarser resolution (equi-finality)

Επιδράσεις Πρώτου και Δεύτερου Βαθμού



1ου βαθμού επιδράσεις

Η χωρική κατανομή μιας μεταβλητής εξηγείται από περιβαλλοντικούς (ή εξωγενείς) παράγοντες. Π.χ., η τιμή μιας μεταβλητής $y(x)$ είναι υψηλή στη θέση x εξαιτίας της τιμής $y'(x)$ μιας άλλης μεταβλητής στην ίδια θέση x , ή εξαιτίας της τιμής $y'(x')$ μιας άλλης μεταβλητής σε μια παρακείμενη θέση x'

2ου βαθμού επιδράσεις

Η χωρική κατανομή μιας μεταβλητής εξηγείται από αλληλεπίδραση (ή ενδογενείς) παράγοντες. Π.χ., η τιμή μιας μεταβλητής $y(x)$ είναι χαμηλή στη θέση x εξαιτίας της τιμής $y(x')$ της ίδιας μεταβλητής σε μια παρακείμενη θέση x' , υπό την προϋπόθεση ότι και οι δύο θέσεις x και x' υπόκεινται στις ίδιες περιβαλλοντικές συνθήκες



Βασικά Σημεία Διάλεξης I

Γεω-χωρικά δεδομένα

- ▶ γεω-αναφερόμενα δεδομένα με τιμές μεταβλητής και συντεταγμένες (η τοπολογία και το πλαίσιο ανάλυσης είναι επίσης σημαντικά)
- ▶ τύποι χωρικών δεδομένων:
 1. δεδομένα χωρικών σημειακών προτύπων – συμβάντα
 2. δεδομένα μεταβλητών με συνεχή χωρική κατανομή – πεδία
 3. δεδομένα επιφανειών ή σε κάρναβο – αντικείμενα
 4. δεδομένα χωρικής αλληλεπίδρασης – ροές

Στόχοι της χωρικής ανάλυσης

- ▶ διερευνητική ανάλυση: ψάχνοντας για χωρικά πρότυπα και σχέσεις μεταξύ μεταβλητών
- ▶ επαγωγική ανάλυση: κατασκευή μοντέλου χωρικής διεργασίας από χωρικά πρότυπα, και εκτίμηση παραμέτρων του μοντέλου

Βασικά Σημεία Διάλεξης II



Χωρική στατιστική

- ▶ μεθοδολογικό πλαίσιο για την ανάλυση χωρικών δεδομένων και την ανάπτυξη στατιστικών μοντέλων, το οποίο λαμβάνει υπ' όψη την χωρική αυτοσυσχέτιση και επιτρέπει την ποσοτικοποίηση της αβεβαιότητας στις προβλέψεις των μοντέλων
- ▶ στην ουσία, η χωρική στατιστική είναι μια μορφή πολυμεταβλητής στατιστικής που αναπτύχθηκε για τη στατιστική ανάλυση χωρικών δεδομένων

Προσοχή

- ▶ κάθε συμπέρασμα που εξάγεται μέσω της χωρικής ανάλυσης είναι άρρηκτα συνδεδεμένο με μια συγκεκριμένη **κλίμακα παρατήρησης**, δηλ. με τις συγκεκριμένες χωρικές μονάδες μέτρησης. Το Modifiable Area Unit Problem και το Ecological Inference Problem αποτελούν απόρροια των παραπάνω
- ▶ τα μοντέλα χωρικών διεργασιών προσπαθούν να επεξηγήσουν χωρικά πρότυπα, και συνήθως περιλαμβάνουν:
 - ▶ επιδράσεις 1ου βαθμού ή περιβαλλοντικές επιρροές
 - ▶ επιδράσεις 2ου βαθμού ή αλληλεπιδράσεις (χωρική αυτοσυσχέτιση)

Ο διαχωρισμός αυτός αφορά χωρικά μοντέλα κι όχι δεδομένα. . .