



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΑΓΡΟΝΟΜΩΝ ΚΑΙ ΤΟΠΟΓΡΑΦΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΓΕΩΓΡΑΦΙΑΣ ΚΑΙ ΠΕΡΙΦΕΡΕΙΑΚΟΥ ΣΧΕΔΙΑΣΜΟΥ

**ΑΝΑΠΤΥΞΗ ΕΚΠΑΙΔΕΥΤΙΚΟΥ ΥΛΙΚΟΥ ΓΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ
ΣΕ ΠΕΡΙΒΑΛΛΟΝ ΓΕΩΓΡΑΦΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΩΝ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ του ΓΕΩΡΓΙΟΥ Σ. ΜΑΡΓΑΡΑ

ΕΠΙΒΛΕΠΩΝ : ΚΟΥΤΣΟΠΟΥΛΟΣ ΚΩΣΤΗΣ Καθηγητής Ε.Μ.Π.

ΦΕΒΡΟΥΑΡΙΟΣ 2010

ΠΕΡΙΕΧΟΜΕΝΑ

Περιεχόμενα	1
Εισαγωγή	3

ΚΕΦΑΛΑΙΟ 1: Βασικές έννοιες της δειγματοληψίας και της στατιστικής

1.1	Γενικά	6
1.2	Πλεονεκτήματα της Δειγματοληψίας	8
1.3	Πηγές Σφαλμάτων στη Δειγματοληψία	9
1.4	Σχέδια Δειγματοληψίας	10
1.5	Η αντιπροσωπευτικότητα του δείγματος	10
1.6	Η δειγματοληψία ως μέρος της γεωγραφικής μελέτης	11
1.7	Χωρική Δειγματοληψία	13
	1.7.1 Βήματα Δειγματοληπτικής διαδικασίας	14
	1.7.2 Δειγματοληπτικό υπόβαθρο	15
	1.7.3 Επιλογή δειγματοληπτικής μεθόδου	16
1.8	Τυχαία δειγματοληψία	17
	1.8.1 Σημειακή	18
	1.8.2 Γραμμική	19
	1.8.3 Επιφανειακή	20
1.9	Συστηματική Δειγματοληψία	20
	1.9.1 Σημειακή	21
	1.9.2 Γραμμική	21
	1.9.3 Επιφανειακή	22
1.10	Στρωματοποιημένη Δειγματοληψία	24
	1.10.1 Στρωματοποιημένη τυχαία	25
	1.10.2 Στρωματοποιημένη συστηματική	25
	1.10.3 Στρωματοποιημένη συστηματική μη γραμμική δειγματοληψία	26
1.11	Εφαρμογή της Κανονικής Κατανομής	27
	1.11.1 Μέτρα ακρίβειας – Διαστήματα εμπιστοσύνης	29
	1.11.2 Εκτιμήσεις διαστήματος εμπιστοσύνης για μετρήσεις	31
	1.11.3 Εκτιμήσεις διαστήματος εμπιστοσύνης για απαριθμήσεις	32
1.12	Μέγεθος του δείγματος και εκτίμηση του	32
1.13	Έλεγχος Υποθέσεων – Test X^2	33

ΚΕΦΑΛΑΙΟ 2 : Γεωγραφικά Συστήματα Πληροφοριών

2.1	Γενικά	36
2.2	Τα μέρη ενός ΓΣΠ	37
2.3	Δομές Δεδομένων	38
2.4	Πηγές Δεδομένων για τα ΓΣΠ	41
2.5	Επεξεργασία δεδομένων – Χωρική ανάλυση	41

ΚΕΦΑΛΑΙΟ 3 : Εφαρμογή

3.1	Γενικά	43
3.2	Αναλυτικά βήματα Εφαρμογής μεθόδων δειγματοληψίας	45
3.3	Εφαρμογή Δοκιμαστικής Δειγματοληψίας	50
3.4	Εύρεση Απαιτούμενου Μεγέθους Δείγματος	55
3.5	Εφαρμογή Απλής Τυχαίας Δειγματοληψίας	57
3.6	Στρωματοποιημένη Τυχαία Δειγματοληψία	63
3.7	Συστηματική Δειγματοληψία	71
3.8	Γραμμική τυχαία δειγματοληψία	75
3.9	Γραμμική συστηματική δειγματοληψία	79
	Παρατηρήσεις	82

ΠΑΡΑΡΤΗΜΑ

	Εγκατάσταση Hawth's Analysis Tools	83
	Το λογισμικό ArcGIS 9.X	84
	Πίνακας Κανονικής κατανομής	86
	Πίνακας Κρίσιμων τιμών του X^2	87
	Πινάκες αποτελεσμάτων δειγματοληψιών	88
	Βιβλιογραφία	92

ΕΙΣΑΓΩΓΗ

Το αντικείμενο της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη εκπαιδευτικού υλικού για εφαρμογή των κυριότερων μεθόδων χωρικής δειγματοληψίας σε περιβάλλον γεωγραφικών συστημάτων πληροφοριών. Το υλικό αυτό αποτελείται από χωρικά δεδομένα και αναλυτικές οδηγίες εφαρμογής των δειγματοληπτικών μεθόδων και μπορεί να αποτελέσει μέρος της σειράς ασκήσεων του υποχρεωτικού προπτυχιακού μαθήματος Γεωγραφία και Ανάλυση Χώρου που διδάσκεται στο 6ο εξάμηνο της σχολής Αγρονόμων και Τοπογράφων Μηχανικών.

Οποιαδήποτε προσέγγιση - επέμβαση στο χώρο συνδέεται άμεσα με την έννοια της γεωγραφικής μελέτης. Συστατικά μέρη κάθε γεωγραφικής μελέτης αποτελούν, ο καθορισμός του προβλήματος, η διαδικασία από στοιχεία σε πληροφορία και τα συμπεράσματα. Η διαδικασία από στοιχεία σε πληροφορία δημιουργεί και συνδυάζει την απαραίτητη πληροφορία για την επίλυση του προβλήματος, όπως αυτό ορίστηκε και οριοθετήθηκε στο πρώτο βήμα της μελέτης, στον καθορισμό του προβλήματος. Τελικά η υλοποίηση του στόχου της μελέτης, οι εναλλακτικές προσεγγίσεις και οι απαντήσεις στα ερωτήματα που τίθενται στο πρώτο βήμα αποτελούν τα συμπεράσματα της μελέτης, τα οποία με τη σειρά τους μπορούν να λειτουργήσουν σαν νέα δεδομένα για επανεξέταση, επέκταση, αναθεώρηση του υπάρχοντος προβλήματος (αναδραστικός μηχανισμός).

Όπως γίνεται αντιληπτό βασικό μέρος κάθε μελέτης είναι η συλλογή των στοιχείων που θα επεξεργαστούν, θα αναλυθούν και τελικά θα αποδώσουν τη χρήσιμη πληροφορία. Τα στοιχεία μπορεί να προέρχονται από υπάρχουσες συλλογές – αρχειακές καταγραφές ή να αποτελούν παρατηρήσεις – μετρήσεις δηλαδή να προέλθουν από εργασίες πεδίου.

Στις περισσότερες περιπτώσεις η δυνατότητα να συλλέξουμε χαρακτηριστικά του συνόλου των στοιχείων που μας ενδιαφέρουν είναι πολύ δύσκολο έως αδύνατο. Με την τεχνική της δειγματοληψίας καθίσταται δυνατό, συλλέγοντας χαρακτηριστικά από μερικά στοιχεία και εξάγοντας συμπεράσματα για τα χαρακτηριστικά του συνόλου των στοιχείων, με ορισμένα όρια αβεβαιότητας.

Κύριος στόχος της εργασίας είναι η παρουσίαση και εφαρμογή των δειγματοληψιών σε έναν χάρτη χρήσεων γης, ώστε λαμβάνοντας δείγματα με τις διάφορες μεθόδους να εκτιμηθούν τα ποσοστά των χρήσεων γης του χάρτη. Η εφαρμογή αναπτύχθηκε με αναλυτικό τρόπο (κατά βήματα) με σκοπό να είναι εύκολα εφαρμόσιμη από σπουδαστές, μιας και αποτελεί εκπαιδευτικό υλικό. Η διαδικασία εφαρμογής τους δεν απαιτεί κάποιες ιδιαίτερες γνώσεις σε περιβάλλον γεωγραφικών συστημάτων πληροφοριών αλλά παράλληλα φέρνει σε επαφή τους σπουδαστές με κάποιες βασικές λειτουργίες των ΓΣΠ.

Προηγούμενες εργασίες σχετικές με το θέμα εφαρμογής δειγματοληπτικών μεθόδων στο πλαίσιο της διδασκαλίας τους, ως βασικές μέθοδοι συλλογής στοιχείων, ήταν οι περισσότερες αναλογικά σχεδιασμένες, δηλαδή χωρίς τη χρήση ΓΣΠ στην εφαρμογή τους αλλά απευθείας στον τυπωμένο χάρτη και μία εξ αυτών ήταν σχεδιασμένη σε ΓΣΠ, στο πακέτο ArcCIS 8, με προβλήματα συμβατότητας στις πιο σύγχρονες εκδόσεις του λογισμικού πακέτου που βρίσκονται εγκατεστημένες στο κέντρο Γεωπληροφορικής της Σ.Α.Τ.Μ.

Στο πρώτο κεφάλαιο της εργασίας αναπτύσσονται βασικές έννοιες της δειγματοληψίας και της επαγωγικής στατιστικής, αναλύονται τα πλεονεκτήματα και τα μειονεκτήματα που παρουσιάζονται από τη χρήση δειγμάτων έναντι της καθολικής απογραφικής μεθόδου και έπειτα αναφέρονται οι κυριότερες χωρικές δειγματοληπτικές μέθοδοι, τα

πλεονεκτήματα – μειονεκτήματά τους, οι τρόποι εφαρμογής τους. Επίσης αναφέρονται οι σχέσεις για τον υπολογισμό του μεγέθους του δείγματος συναρτήσει του επιλεγόμενου επιπέδου εμπιστοσύνης και περιθωρίου σφάλματος.

Στο δεύτερο κεφάλαιο δίνονται ορισμοί, χαρακτηριστικά και βασικές λειτουργίες των γεωγραφικών συστημάτων πληροφοριών.

Στο τρίτο κεφάλαιο αναπτύσσεται αναλυτικά η εφαρμογή των δειγματοληπτικών μεθόδων, δίνονται στοιχεία για τα θεματικά επίπεδα του χάρτη – υπόβαθρο και παρουσιάζονται οι λειτουργίες των διαφόρων εργαλείων που χρησιμοποιούνται.

Για την εκτίμηση του μεγέθους του δείγματος και την εξαγωγή κάποιων στατιστικών συμπερασμάτων οι απαραίτητοι υπολογισμοί έχουν γίνει σε φύλλο excel, έτσι ώστε να δίνεται η δυνατότητα αλλαγής των παραμέτρων υπολογισμού του μεγέθους του δείγματος βάσει των απαιτήσεων ακρίβειας. Συγκεκριμένα υπολογίζονται τα διαστήματα εμπιστοσύνης και το τεστ χ^2 .

Η εργασία ολοκληρώνεται με τη παράθεση παρατηρήσεων που προέκυψαν από την εφαρμογή των δειγματοληψιών.

Θεωρώ υποχρέωση μου να ευχαριστήσω τον επιβλέποντα καθηγητή, Κωστή Κουτσόπουλο και τον Δρ Θωμά Χατζηχρήστο, για τη βοήθεια τους στην εκπόνηση της εργασίας.

EDUCATIONAL TOOL FOR SPATIAL SAMPLING IN GEOGRAPHIC INFORMATION SYSTEMS (G.I.S.)

ABSTRACT

The objective of the present dissertation is the development of an educational tool for applying the main methods of spatial sampling using Geographic Information Systems GIS. This tool is comprised by spatial data and detailed directions of sampling methods' application.

Gathering data, analyzing and processing them is the significant part of every study that will attribute to the useful information. These data may originate from already existent compilation and records or from field project. In most cases, the potential of gleaning characteristics of the total of the data that we are interested in, it is not so easy. By using the sampling method is being possible to draw conclusions for the statistic population, within uncertainty limits.

The aim of the dissertation is the presentation and the application of sampling in a land use map using different sampling methods to estimate the rates of land uses. The application was integrated in an analytical way (by steps) in order to facilitate students. Despite the fact that the application processes do not require technical knowledge of Geographical Information Systems, the students are introduced to the basic functions of GIS.

The first chapter of the dissertation presents the basic concept of sampling and inferential statistics. The pros and cons of sampling towards board census methodology are examined. The main spatial sampling methods and their application methods are mentioned. Finally, the mathematic relations for the calculation of the size of the sample in relation to the chosen confidence level and error are presented.

At the second chapter, terms, characteristics and basic concepts of Geographical Information Systems are being introduced.

Finally, at the last chapter, the application of sampling methods (simple random point sampling, systematic point sampling, stratifying point sampling, random transect sampling and systematic transect sampling) is studied in detail. Information is given for the layers of the map and the functions of the tools used are presented. In order to estimate the size of the sample and to draw some statistical conclusions, the necessary calculations were made in EXCEL, which allows the introduction of calculation parameters of the size of the sample based on the requirements of precision set.

The dissertation is completed by displaying some comments and remarks on the sampling applications.

ΚΕΦΑΛΑΙΟ 1: Βασικές έννοιες της Δειγματοληψίας και της Στατιστικής

1.1 Γενικά

Στατιστική είναι η επιστήμη η οποία ασχολείται με τη συλλογή, τη ταξινόμηση, την ανάλυση και την ερμηνεία αριθμητικών γεγονότων και στοιχείων.

Η επιστήμη της στατιστικής είναι ένα σύνολο αρχών και μεθοδολογιών για:

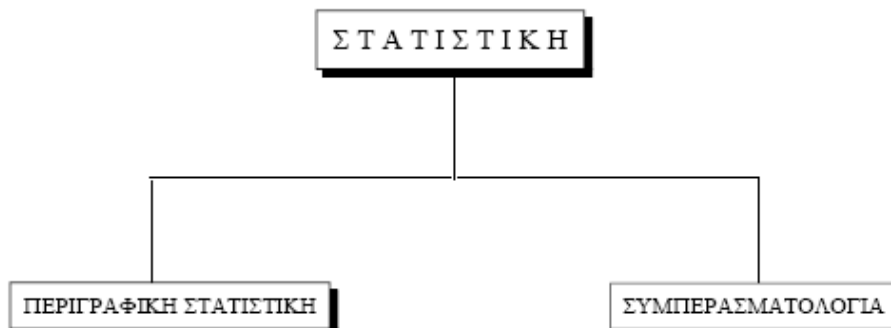
- Το σχεδιασμό της διαδικασίας συλλογής των δεδομένων.
- Σύνοψη, περιγραφή και εξέταση των διαφόρων στοιχείων.
- Τη χρήση δείγματος από στοιχεία και την εξαγωγή συμπερασμάτων για το σύνολο των στοιχείων από τα οποία προέρχονται τα δείγματα.

Ο κλάδος της Στατιστικής που ασχολείται με τον πρώτο στόχο λέγεται **σχεδιασμός πειραμάτων** (experimental design).

Στη δεύτερη περίπτωση αναφερόμαστε στον κλάδο της στατιστικής που ονομάζεται **περιγραφική στατιστική** (descriptive statistics).

Ενώ ο τρίτος τύπος προβλημάτων αναφέρεται και ως **επαγωγική στατιστική** ή **στατιστική συμπερασματολογία** (inferential statistics).

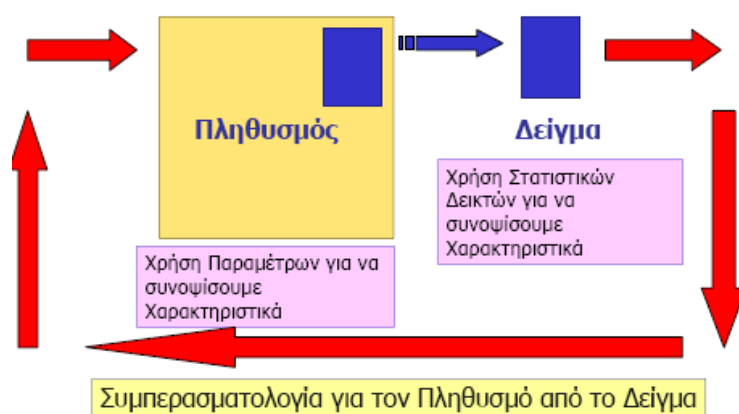
Στη στατιστική ορολογία, το σύνολο των στοιχείων που θέλουμε να περιγράψουμε και να μελετήσουμε ως προς μια συγκεκριμένη ιδιότητα ή χαρακτηριστικό ονομάζεται **πληθυσμός**. Κάθε πεπερασμένο υποσύνολο του πληθυσμού ονομάζεται **δείγμα**. Όπως υπονοείται και από τη λέξη υποσύνολο τα μέλη του δείγματος αποτελούν και μέλη του πληθυσμού. Το χαρακτηριστικό-ά ή οι ιδιότητες των στοιχείων που μελετώνται ονομάζονται και **μεταβλητές**.



Η στατιστική σήμερα χρησιμοποιείται ευρύτατα σε όλους σχεδόν τους τομείς της ανθρώπινης δραστηριότητας. Βασικές έννοιες της στατιστικής έχουν εισχωρήσει και ενσωματωθεί σε όλες τις επιστήμες. Από τις Ανθρωπιστικές, Νομικές και Κοινωνικές επιστήμες (Αρχαιολογία, Κοινωνιολογία, Δημογραφία κ.α.), τις Φυσικές επιστήμες, τις επιστήμες Υγείας, τις Τεχνολογικές και φυσικά τις επιστήμες Οικονομίας και Διοίκησης. Η ανάλυση στατιστικών ερευνών είναι το κυριότερο εργαλείο έρευνας σε ένα μεγάλο φάσμα εφαρμογών των παραπάνω επιστημών. Οι έρευνες των ανθρώπινων πληθυσμών – συχνά αναφερόμενες και ως δημοσκοπήσεις – αποτελούν σπουδαίες πηγές βασικής γνώσης των κοινωνικών επιστημών.

Ένα παράδειγμα επαγωγικής στατιστικής: σε δείγμα 100 ανθρώπων μιας πόλης 30.000 κατοίκων ερωτάται τι πρόκειται να ψηφίσουν στις ερχόμενες εκλογές. Αρχικά πρέπει να διασαφηνιστούν κάποιες βασικές έννοιες:

- Με ποιον τρόπο επιλέχτηκε το δείγμα. (πχ. απλή τυχαία δειγματοληψία - θα αναφερθούμε σε επόμενη παράγραφο)
- Γιατί επιλέχτηκε τέτοιο μέγεθος δείγματος (100 ατόμων)
- Ποιος είναι ο πληθυσμός από τον οποίο επιλέχτηκε το δείγμα και στον οποίο θα αναφέρονται τα αποτελέσματα του δείγματος. (οι μόνιμοι κάτοικοι της πόλης που έχουν δικαίωμα ψήφου;)
- Η μεταβλητή που μας ενδιαφέρει (προφανώς η πρόθεση ψήφου).
- Επιμέρους μεταβλητές που θα μπορούσαν να μας ενδιαφέρουν (πρόθεση ψήφου ανά ηλικιακή ομάδα) κτλ.
- Ακρίβεια των αποτελεσμάτων.



Ένας τρόπος για να λάβουμε τις απαραίτητες πληροφορίες που χρειαζόμαστε για κάποιο πληθυσμό είναι να εξετάσουμε όλα τα άτομα (στοιχεία) του πληθυσμού ως προς το χαρακτηριστικό που μας ενδιαφέρει. Η μέθοδος αυτή συλλογής των δεδομένων καλείται **απογραφή** (census). Για παράδειγμα η στατιστική υπηρεσία της χώρας μας (ΕΣΥΕ) κάνει κάθε 10 χρόνια απογραφή του πληθυσμού, η οποία αποτελεί κύρια πηγή δεδομένων δημογραφικού, οικονομικού, εμπορικού και βιομηχανικού χαρακτήρα. Σε πολλές περιπτώσεις η εξέταση όλων των μονάδων του πληθυσμού είναι δύσκολη ή ακόμα και αδύνατη. Ο ερευνητής λοιπόν ‘μαζεύει’ πληροφορίες από κάποια μικρή ομάδα ή υποσύνολο του πληθυσμού, το **δείγμα**. Κάνει λοιπόν τις παρατηρήσεις του στο δείγμα ώστε να γενικεύσει τα συμπεράσματα του για ολόκληρο τον πληθυσμό. Οι αρχές και οι

μέθοδοι για τη συλλογή και ανάλυση δεδομένων από πεπερασμένους πληθυσμούς είναι αντικείμενο της **Δειγματοληψίας** (sampling), που αποτελεί τη βάση της Στατιστικής. Η οργάνωση της συλλογής και επεξεργασίας των σχετικών δεδομένων και πληροφοριών γίνεται κατά τρόπο που για δεδομένη ακρίβεια να επιτυγχάνεται το χαμηλότερο δυνατό κόστος ή, αντιστρόφως, να εξασφαλίζεται η μέγιστη δυνατή ακρίβεια την οποία επιτρέπουν τα μέσα που διαθέτουμε.

1.2 Πλεονεκτήματα της Δειγματοληψίας

Ο βασικός σκοπός της επαγωγικής στατιστικής είναι η εξαγωγή συμπερασμάτων για διάφορα χαρακτηριστικά του στατιστικού πληθυσμού, βασισμένα σε πληροφορία που λαμβάνεται από ένα δείγμα.

Θεωρητικοί και πρακτικοί λόγοι δίνουν προβάδισμα στην δειγματοληψία έναντι της απαρίθμησης και της απογραφής ολόκληρου του πληθυσμού. Πλεονεκτήματα της συνοψίζονται παρακάτω

Η δειγματοληψία είναι απαραίτητη σε πολλά προβλήματα γεωγραφικής έρευνας. Εάν ο πληθυσμός που μελετάται είναι υπερβολικά μεγάλος (ή ακόμα και – θεωρητικά – άπειρος) το να επιτευχθεί μια πλήρης απαρίθμηση – απογραφή δεν είναι δυνατό. Για παράδειγμα η δειγματοληψία είναι η μόνη εναλλακτική όταν μελετώνται οι λόγοι για τους οποίους οικογένειες μιας χώρας μετεγκαθίστανται σε άλλες περιοχές, αφού δεν είναι δυνατό ο μελετητής να έρθει σε επαφή με όλες τις οικογένειες που άλλαξαν τόπο κατοικίας, έστω και σε ένα ορισμένο διάστημα. Επίσης ένας γεωγράφος που αναλύει τις παγκόσμιες αλλαγές στη φύση και την έκταση περιοχών που συμβαίνουν τροπικές βροχές, θα λάβει δείγματα από αυτές τις περιοχές μιας και είναι αδύνατο να έχει στοιχεία για το σύνολο των θέσεων αυτών των περιοχών που πρακτικά είναι άπειρες.

Η δειγματοληψία είναι αποτελεσματική και οικονομική μέθοδος συλλογής πληροφοριών. Με τη βοήθεια ενός δείγματος, μία κατάλληλη ποσότητα δεδομένων που αφορούν τον πληθυσμό μπορεί γρήγορα να αποκτηθεί και να αναλυθεί. Η δειγματοληψία λοιπόν μειώνει τα έξοδα και τον χρόνο και ελαχιστοποιεί μια σειρά από λογιστικά προβλήματα, σε αντίθεση με μια εξέταση του συνόλου του πληθυσμού.

Η δειγματοληψία μπορεί να παρέχει υψηλής λεπτομέρειας πληροφορία. Σε γεωγραφικά προβλήματα στα οποία, η σε βάθος ανάλυση είναι απαραίτητη, μόνο ένας μικρός αριθμός ατόμων ή τοποθεσιών μπορεί να συμπεριληφθεί στη μελέτη.

Αυτά τα λίγα στοιχεία του δείγματος μπορούν έπειτα να εξεταστούν εξονυχιστικά σε συνδυασμό με μια σειρά από πρόσθετες πληροφορίες για κάθε μέλος του δείγματος.

Για παράδειγμα μία μελέτη για την καταναλωτική συμπεριφορά θα απαιτούσε πολυάριθμες ερωτήσεις σε κάθε μέλος του δείγματος όπως για παράδειγμα πόσες φορές την εβδομάδα πηγαίνετε για ψώνια, ποιες περιοχές επισκέπτεστε για ψώνια; κτλ.

Η δειγματοληψία επιτρέπει την επανάληψη της συλλογής πληροφορίας γρήγορα και με μικρό κόστος. Πολλές γεωγραφικές έρευνες απαιτούν την πρόσθετη συλλογή πληροφοριών μετά από κάποιο χρονικό διάστημα, ώστε να εξεταστούν οι όποιες διαφορές προέκυψαν. Για παράδειγμα ένας αστικός γεωγράφος που μελετά την ανάπτυξη μιας περιοχής θα ήθελε να επικεντρωθεί στις απόψεις των κατοίκων πριν και μετά τη κατασκευή ενός μεγάλου εμπορικού κέντρου στην περιοχή τους. Μελετώντας δυναμικά και όχι στατικά τέτοια προβλήματα, η μοναδική λύση είναι η δειγματοληψία ώστε τα

έξοδα σε χρόνο και χρήματα να μην είναι υπέρογκα. Η δειγματοληψία συχνά παρέχει αποτελέσματα υψηλής ακρίβειας, αρκετές φορές τα αποτελέσματα της είναι πιο ακριβή από μια απογραφή η οποία απαιτεί πολυάριθμο προσωπικό για την διεξαγωγή της, προσωπικό που δεν είναι δυνατό να είναι πάντοτε εκπαιδευμένο και εξειδικευμένο.

1.3 Πηγές Σφαλμάτων στη Δειγματοληψία

Όπως προαναφέρθηκε, κύριος σκοπός της δειγματοληψίας είναι η εξαγωγή αξιόπιστων συμπερασμάτων για τα χαρακτηριστικά ενός πληθυσμού με τη μελέτη ενός μικρού μέρους των μελών του πληθυσμού.

Για την διασφάλιση αμερόληπτου και αντιπροσωπευτικού δείγματος – έννοιες που θα αναλυθούν περαιτέρω σε επόμενες παραγράφους - είναι απαραίτητο στη διαδικασία να εμπεριέχεται το στοιχείο της τυχαιότητας.

Ωστόσο, η τυχαιότητα από μόνη της δεν εξασφαλίζει την αμεροληψία του δείγματος, υπάρχουν και αρκετές άλλες πιθανές πηγές σφαλμάτων.

Δυστυχώς, συχνά δε γνωρίζουμε εάν το δείγμα μας είναι απολύτως αντιπροσωπευτικό. Αυτό συμβαίνει διότι δεν είναι δυνατό να είναι γνωστές όλες οι παράμετροι του πληθυσμού από όπου προέρχεται το δείγμα. Οι όποιες αβεβαιότητες υπάρχουν συνδέονται άμεσα με το δείγμα και δημιουργούν σφάλματα σε αυτό.

Οι μετρήσεις ακρίβειας και ορθότητας βοηθούν ώστε να κατηγοριοποιηθούν οι πηγές των σφαλμάτων. Τα αποτελέσματα από ένα μικρό δείγμα είναι πιθανό να μην είναι ακριβή και ορθά. Σε αυτή τη περίπτωση η ακρίβεια μπορεί να αυξηθεί, αυξάνοντας το μέγεθος του δείγματος. Μεγαλύτερο δείγμα από την άλλη μεριά δεν παρουσιάζει μεγάλες διαφορές μεταξύ των στοιχείων του, είναι πιο δαπανηρό και απαιτεί μεγαλύτερο χρόνο για τη συλλογή του. Έτσι θα πρέπει να επιλέγεται μια ισορροπία ανάμεσα στην ακρίβεια του δείγματος και στη προσπάθεια που απαιτείται για την απόκτηση του.

Οι ανακρίβειες που παρουσιάζει ένα δείγμα είναι ένα αρκετά περίπλοκο ζήτημα. Συστηματική μεροληπτικότητα (Systematic bias) μπορεί να εισαχθεί στη δειγματοληπτική διαδικασία με πολλούς διαφορετικούς τρόπους. Αρκετές ανακρίβειες προέρχονται από την ίδια τη διαδικασία που ακολουθείται. Για παράδειγμα τα στοιχεία ή τα μέλη του πληθυσμού μπορεί να μην ταιριάζουν σωστά με τα στοιχεία ή τα μέλη του δείγματος. Ένας αστικός γεωγράφος που μελετά τη χωροθέτηση των νέων οικοδομών μιας περιοχής θα μπορούσε να χρησιμοποιήσει ως πηγή για τα δεδομένα του τη λίστα οικοδομικών αδειών. Αν η λίστα αυτή περιέχει τις ανακαινίσεις οικοδομών ως νέες οικοδομές, κάποιες από αυτές θα χρησιμοποιηθούν αναπόφευκτα στο δείγμα ως νέες οικοδομές.

Η επιλογή εσφαλμένης μεθόδου είναι επίσης μία σημαντική πηγή για ανακριβή αποτελέσματα. Η πολυπλοκότητα ενός γεωγραφικού προβλήματος απαιτεί συνήθως και πιο εξεζητημένο σχέδιο από την απλή τυχαιά δειγματοληψία.

Σε άλλα γεωγραφικά προβλήματα μια ακατάλληλη για την περίσταση μέθοδος συλλογής δεδομένων μπορεί να δημιουργήσει σφάλματα και έτσι να μην απεικονίζεται η πραγματικότητα. Για παράδειγμα ερωτηματολόγιο δια αλληλογραφίας αντί τηλεφωνικής συνέντευξης.

Άλλες ανακρίβειες μπορεί να προέρχονται από λογιστικά ή σφάλματα που προέρχονται από το προσωπικό και δε σχετίζονται άμεσα με αυτή καθεαυτή τη διαδικασία της

δειγματοληψίας. Ανακολουθίες στη συλλογή των δεδομένων ή στις συνεντεύξεις επηρεάζουν σημαντικά τα αποτελέσματα. Λάθη μπορεί να προκύψουν κατά την εγγραφή και την πινακοποίηση της πληροφορίας.

Ο μελετητής λοιπόν θα πρέπει να αναπτύσσει προσεκτικά την διαδικασία της δειγματοληψίας ώστε να περιορίζονται τα σφάλματα σε ένα αποδεκτό επίπεδο.

Οι περισσότερες πηγές σφαλμάτων μπορούν να περιοριστούν ή ακόμα και να εξαλειφθούν εάν όλα τα βήματα της διαδικασίας σχεδιαστούν προσεκτικά.

1.4 Σχέδια Δειγματοληψίας

Τα σχέδια δειγματοληψίας γενικά κατανέμονται σε δύο κατηγορίες:

- Τυχαία Δειγματοληψία
- Μη Τυχαία Δειγματοληψία

Η Τυχαία δειγματοληψία (Random Sampling) συνήθως αναφέρεται σαν δειγματοληψία πιθανότητας (Probability Sampling), διότι εάν η δειγματοληπτική διαδικασία είναι τυχαία, οι νόμοι πιθανότητας μπορούν να εφαρμοστούν, έτσι το υπόδειγμα της κατανομής δειγματοληψίας είναι διαθέσιμο.

Ο όρος τυχαίο δείγμα (Random Sample) δεν χρησιμοποιείται για να περιγράψει τα δεδομένα του δείγματος, αλλά τη διαδικασία που εφαρμόστηκε για την επιλογή του δείγματος. Η τυχαιότητα λοιπόν είναι μια ιδιότητα που χαρακτηρίζει την δειγματοληπτική διαδικασία και όχι ένα συγκεκριμένο δείγμα.

Μη Τυχαία δειγματοληψία (Non-random Sampling) είναι η διαδικασία επιλογής ενός δείγματος, χωρίς τη χρήση της "τυχαιότητας". Ένα μη τυχαίο δείγμα επιλέγεται με βάση άλλες θεωρήσεις, όπως για παράδειγμα την "εμπειρική κρίση". Το σπουδαιότερο χαρακτηριστικό της μη τυχαίας δειγματοληψίας είναι ότι υπόκειται σε δειγματοληπτική μεταβλητότητα (διακυμάνσεις της δειγματοληψίας) και δεν υπάρχει τρόπος να γνωρίζουμε το υπόδειγμα της μεταβλητότητας κατά την διαδικασία.

Στη παρούσα εργασία γίνεται αναφορά και μελέτη της τυχαίας δειγματοληψία και των διαφόρων μεθόδων της (Types of Probability Sampling).

1.5 Η αντιπροσωπευτικότητα του δείγματος

Η παράθεση της δειγματοληπτικής μεθοδολογίας αποσκοπεί στον προσδιορισμό ενός δείγματος με το μικρότερο, κατά το δυνατό, μέγεθος, ώστε να μειωθεί αναλόγως το κόστος της δειγματοληψίας, με τη βασική προϋπόθεση ότι το δείγμα θα αναπαράγει με τον καλύτερο δυνατό τρόπο τα χαρακτηριστικά του αρχικού πληθυσμού από τον οποίο

επιλέχτηκε, παράλληλα οι όποιες διαφοροποιήσεις του πληθυσμού να εκφράζονται όσο το δυνατόν πιο πιστά.

Δυστυχώς, η εκλογή αντιπροσωπευτικού δείγματος, παρ' ότι είναι επιθυμητή, είναι σχεδόν αδύνατη στις περισσότερες περιπτώσεις. Το γεγονός αυτό πιστοποιείται με τα ακόλουθα δεδομένα. Εάν π.χ. η χαρακτηριστική ιδιότητα ενός πληθυσμού έχει τις τιμές :

$$X_1, X_2, \dots, X_r$$

με αντίστοιχες απόλυτες συχνότητες :

$$N_1, N_2, \dots, N_r \quad \text{όπου} \quad \sum_{i=1}^r N_i = N$$

τότε για ένα αντιπροσωπευτικό δείγμα οι αντίστοιχες απόλυτες συχνότητες θα πρέπει να είναι :

$$v_1, v_2, \dots, v_r \quad \text{όπου} \quad v_i = v \frac{N_i}{N}, \quad v = \sum_{i=1}^r v_i$$

Αυτό σημαίνει ότι για τον προσδιορισμό ενός αντιπροσωπευτικού δείγματος θα πρέπει να είναι γνωστές οι απόλυτες συχνότητες N_i ($i = 1, 2, \dots, r$) του αρχικού πληθυσμού.

Όταν όμως οι πληροφορίες για την κατανομή του αρχικού πληθυσμού είναι δεν είναι τόσο ακριβείς, παύει να έχει νόημα η εφαρμογή δειγματοληπτικών μεθόδων.

Μπορεί να αναφερθεί, ότι μια διέξοδος στο προβληματισμό αυτό, της λήψης ενός καλού δείγματος, προσφέρει η εκλογή που στηρίζεται στα αποτελέσματα πειραμάτων τύχης, των οποίων η μεθοδολογία καθορίζεται αυστηρά με την επαναληπτική εκτέλεση μεγάλου πλήθους αυτών των ίδιων των πειραμάτων. Έτσι η αξιολόγηση των αποτελεσμάτων αποτελεί αντικείμενο της επαγωγικής στατιστικής, όπου η εξαγωγή συμπερασμάτων για τις ιδιότητες του αρχικού πληθυσμού, κυρίως για τον τρόπο κατανομής του διερευνώμενου χαρακτηριστικού, στηρίζεται στα αποτελέσματα της τυχαίας δειγματοληψίας.

1.6 Η δειγματοληψία ως μέρος της γεωγραφικής μελέτης

Με το όρο γεωγραφική μελέτη εννοούμε μια μελέτη που αναλύει τη δομή του χώρου, τις σχέσεις και αλληλεξαρτήσεις των στοιχείων του και τις διαδικασίες αλλαγής του. Αναπόσπαστο κομμάτι μιας γεωγραφικής μελέτης είναι η συλλογή στοιχείων, από την επεξεργασία των οποίων θα προκύψει η πληροφορία (αναφέρεται και ως διαδικασία από στοιχεία σε πληροφορία).

Τα στοιχεία μπορούν να διαχωριστούν σε δύο βασικές κατηγορίες:

- Ποσοτικά στοιχεία (quantitative)
- Ποιοτικά ή κατηγορικά (qualitative – categorical)

Η διάκριση τους σε θεωρητικό επίπεδο είναι προφανής, τα μεν ποσοτικά αναφέρονται σε μεγέθη που μπορούν να μετρηθούν σε κάποια αριθμητική κλίμακα, σε αντίθεση με τα ποιοτικά ή κατηγορικά τα οποία δεν εκφράζουν κάποια ποσότητα αλλά κάποια ιδιότητα και συνήθως ταξινομούνται.

Στη πράξη η συγκεκριμένη διάκριση είναι αρκετές φορές προβληματική, άλλωστε αυτό που ενδιαφέρει περισσότερο κατά τη συλλογή στοιχείων για κάποια γεωγραφική μελέτη είναι το *επίπεδο μέτρησης* των στοιχείων. Όταν λοιπόν αναφερόμαστε σε επίπεδο μέτρησης στοιχείων στην ουσία αναφερόμαστε σε κάποιο από τα 4 επίπεδα:

- **Ποιοτικό:** ταξινομεί τα αντικείμενα σε κατηγορίες,
- **Ιεραρχικό:** εκφράζει σύγκριση ανάμεσα σε κατηγορίες π.χ. μικρότερου – μεγαλύτερου,
- **Ποσοτικό:** εκφράζει ποσοτικά μια ιεραρχική σχέση,
- **Αναλογικό:** έχει το πλεονέκτημα μιας φυσικής αφετηρίας (σταθερό μηδέν).

Από τη στιγμή που έχουν προσδιοριστεί το περιεχόμενο, οι πηγές άντλησης και το επίπεδο μέτρησης των στοιχείων, θα πρέπει να αποφασιστεί η μέθοδος συλλογής τους.

Αυτή μπορεί να είναι *απογραφική* (καθολική - πλήρης) ή *δειγματοληπτική* (δείγμα - κομμάτι του πληθυσμού).

1.7 Χωρική Δειγματοληψία

Στη χωρική δειγματοληψία, η χωρική αναφορά του κάθε στοιχείου, αντικειμένου είναι μείζονος σημασίας. Η αναφορά αυτή μπορεί να εκφράζεται με τη βοήθεια ενός συστήματος αναφοράς, που αποδίδει μοναδικές συντεταγμένες σε κάθε σημείο στο χώρο. Για παράδειγμα χωρική δειγματοληψία χρησιμοποιείται όταν έχουμε ένα χάρτη από συνεχώς κατανεμημένες μεταβλητές (όπως το είδος βλάστησης, ο τύπος του εδάφους ή το pH των υδάτινων επιφανειών) και γίνεται επιλογή ενός δείγματος από τον χάρτη. Δείγματα πληθυσμών χρησιμοποιούνται είτε για την αποκόμιση μιας γενικής εντύπωσης της περιοχής ενδιαφέροντος, είτε για να εκτιμηθούν ορισμένα χαρακτηριστικά της. Στη στατιστική χρήση των δειγμάτων για την εκτίμηση χαρακτηριστικών της περιοχής που μελετάται εστιάζει κυρίως η μοντέρνα γεωγραφία.

Βασικές έννοιες και παράμετροι κάθε χωρικής δειγματοληπτικής διαδικασίας αποτελούν και εδώ:

- Ο Πληθυσμός
- Το Δείγμα

Με την έννοια του **πληθυσμού** ορίζεται ένα σύνολο στοιχείων που πρόκειται να μελετηθεί, ως προς μία ή περισσότερες ιδιότητες – χαρακτηριστικά (μεταβλητές). Θα πρέπει να διευκρινιστεί ιδιαίτερα ότι τα στοιχεία του πληθυσμού μπορεί να είναι οποιοδήποτε σύνολο αντικειμένων ή τιμών που προέκυψαν από πειράματα ή παρατηρήσεις.

Για να γίνει πιο κατανοητή η έννοια του πληθυσμού, ας υποθεθεί ότι υπάρχει ανάγκη να εξακριβωθούν διάφορα χαρακτηριστικά μιας περιοχής, για παράδειγμα το μέγεθος των αγροτεμαχίων και το είδος της καλλιέργειας τους. Η καταγραφή καθενός από αυτά τα χαρακτηριστικά δημιουργεί δύο πληθυσμούς. Ο ένας θα αποτελείται από το σύνολο των **μετρήσεων** (measurements) για το μέγεθος όλων των αγροτεμαχίων της περιοχής. Ο δεύτερος θα αποτελείται από το σύνολο των **απαριθμήσεων** (counts) για τα είδη των καλλιεργειών στην περιοχή.

Πληθυσμός, λοιπόν στη στατιστική νοείται ένα σύνολο από μετρήσεις ή απαριθμήσεις που προέρχονται από όλα τα αντικείμενα που έχουν ένα ή περισσότερα κοινά χαρακτηριστικά. Στο παραπάνω παράδειγμα, το κοινό χαρακτηριστικό αναφερόταν στο αντικείμενο ‘‘αγροτεμάχια μιας ορισμένης περιοχής’’.

Στη στατιστική, **δείγμα** θεωρείται κάθε πεπερασμένο υποσύνολο του πληθυσμού.

Σε κάθε περίπτωση ο καλύτερος τρόπος για την μελέτη των ιδιοτήτων ενός γεωγραφικού πληθυσμού είναι η μελέτη κάθε αντικειμένου χωριστά. Τις περισσότερες φορές βέβαια δεν είναι αναγκαίο, ούτε οικονομικό, να αναζητούνται μετρήσεις ή απαριθμήσεις από ολόκληρο τον πληθυσμό. Έτσι χρησιμοποιείται ένα μέρος - δείγμα του πληθυσμού. Ένα δείγμα επομένως αντικαθιστά τον πληθυσμό για να εξοικονομηθεί χρόνος, χρήματα και προσπάθεια ή σε περίπτωση που οι μετρήσεις σε ολόκληρο τον πληθυσμό δεν είναι δυνατό ή πρακτικό να γίνουν.

Για να είναι αξιόπιστη η αντικατάσταση του πληθυσμού από το δείγμα θα πρέπει το δεύτερο να είναι **αντιπροσωπευτικό**, δηλαδή να εκφράζει τις διαφορές του πληθυσμού. Για να είναι αξιόλογο ένα δείγμα θα πρέπει επίσης να είναι **αμερόληπτο** (unbiased sample). Η μεροληπτικότητα είναι αποτέλεσμα της επιλογής μη αντιπροσωπευτικών τμημάτων του πληθυσμού. Τα δείγματα δεν θα πρέπει να επιλέγονται με

υποκειμενικότητα από τον ερευνητή. Ο ευκολότερος τρόπος λήψης ενός καλού δείγματος – που να αναπαράγει τον αρχικό πληθυσμό κατά τον καλύτερο τρόπο – στηρίζεται στα αποτελέσματα πειραμάτων τύχης. Άλλωστε, σκοπός της δειγματοληψίας δεν είναι να μετρήσουμε τις χαρακτηριστικές ιδιότητες του δείγματος, αλλά απ' αυτό να λάβουμε εκτιμήσεις των πραγματικών μεγεθών του πληθυσμού.

1.7.1 Βήματα Δειγματοληπτικής διαδικασίας

Σε γεωγραφικές έρευνες στις οποίες συλλέγεται πληροφορία από δείγματα, η διαδικασία περιλαμβάνει μια σειρά από βήματα που αναφέρονται στη παρούσα παράγραφο.

Βήμα 1^ο: Καθορισμός του πληθυσμού αναφοράς και της περιοχής αναφοράς.

- *πληθυσμός αναφοράς:* Το σύνολο των μελών από όπου συλλέγεται η πληροφορία
- *περιοχή αναφοράς:* Ολόκληρη η περιοχή ή το σύνολο των θέσεων από τις οποίες συλλέγεται η πληροφορία

Βήμα 2^ο: Προσδιορισμός του δειγματοληπτικού πληθυσμού και της περιοχής δειγματοληψίας από το δειγματοληπτικό υπόβαθρο.

- *δειγματοληπτικός πληθυσμός:* Το σύνολο των μελών από τα οποία στη πραγματικότητα επιλέγεται το δείγμα.
- *δειγματοληπτική περιοχή:* Η περιοχή ή το σύνολο των θέσεων όπου βρίσκονται τα στοιχεία που λαμβάνουν μέρος στη διαδικασία.

Βήμα 3^ο: Επιλογή δειγματοληπτικής μεθόδου:

- Η τυχαία δειγματοληψία (probability sampling) προτιμάται έναντι της μη τυχαίας (non-probability sampling)
- Τα κυριότερα είδη τυχαίας δειγματοληψίας είναι: η απλή τυχαία, η συστηματική, η στρωματοποιημένη.

Βήμα 4^ο: Οι μέθοδοι της συλλογής δεδομένων περιλαμβάνουν: άμεσες παρατηρήσεις, μετρήσεις πεδίου, αποστολή ερωτηματολογίων, προσωπικές συνεντεύξεις, τηλεφωνικές συνεντεύξεις.

Βήμα 5^ο: Διεξαγωγή προκαταρκτικού τεστ και έλεγχος της λειτουργίας της μεθόδου που επιλέχτηκε, διόρθωση των όποιων προβλημάτων – που μπορούν να οδηγήσουν σε δειγματοληπτικά σφάλματα – και τελική απόφαση για το μέγεθος του δείγματος.

Βήμα 6^ο: Συλλογή δείγματος:

- Η συνοχή στη μέθοδο συλλογής του δείγματος και στη διαδικασία (procedure) είναι μείζονος σημασίας.
- Εξασφάλιση υψηλού επιπέδου αποτελεσμάτων.

1.7.2 Δειγματοληπτικό υπόβαθρο

Το δειγματοληπτικό υπόβαθρο αποτελεί τον στατιστικό πληθυσμό από όπου λαμβάνεται το δείγμα και μπορεί να είναι χωρικό ή μη χωρικό.

Όταν η χωρική διάσταση των στοιχείων, αντικειμένων, φαινομένων που μελετώνται και κατ'επέκταση η θέση τους, είναι καθοριστικής σημασίας, είναι απαραίτητο το δείγμα να ληφθεί από ένα χωρικό δειγματοληπτικό υπόβαθρο. Τέτοιο υπόβαθρο μπορεί να είναι ένας χάρτης στον οποίο παρουσιάζεται η μεταβλητή που εξετάζουμε.

Στη περίπτωση που η χωρική διάσταση ενός πληθυσμού στοιχείων, αντικειμένων, φαινομένων είναι μεν υπαρκτή αλλά όχι και καθοριστική, το υπόβαθρο μπορεί να είναι πλέον μη χωρικό, δηλαδή ένας κατάλογος – από τα μέλη του οποίου επιλέγεται το δείγμα. Φυσικά, τα μέλη του καταλόγου ως σύνολο, δηλαδή ο στατιστικός πληθυσμός, θα εντοπίζονται στο χώρο του προβλήματος, χωρίς όμως να είναι γνωστή η θέση τους.

Σε ένα χωρικό δειγματοληπτικό υπόβαθρο, η θέση είναι ένα βασικό τμήμα της μεταβλητότητας καθεμιάς από τις παρατηρήσεις και για αυτό το λόγο πρέπει να αντιπροσωπεύει το συνολικό φάσμα της χωρικής μεταβλητότητας. (spatial variability). Αν για παράδειγμα πρέπει να υπολογιστεί το ποσοστό των χρήσεων της γης ανά κατηγορία σε μια περιοχή που απεικονίζεται σε ένα χάρτη, λαμβάνοντας ένα δείγμα από τον χάρτη της περιοχής, θα πρέπει όλα τα τμήματα του χάρτη να αντιπροσωπεύονται στο δειγματοληπτικό υπόβαθρο. Μη χωρικά υπόβαθρα είναι αποδεκτά στη περίπτωση που το ενδιαφέρον εστιάζεται σε ένα χαρακτηριστικό μιας περιοχής ως σύνολο.

Στη περίπτωση των χωρικών υποβάθρων πρέπει να καθοριστεί το είδος των παρατηρήσεων δηλαδή να οριστεί η χωρική δειγματοληπτική μονάδα. Βασικά υπάρχουν τρεις τέτοιες μονάδες, **σημεία, γραμμές, επιφάνειες καννάβων**. Έτσι έχουμε σημειακές δειγματοληψίες, γραμμικές και δειγματοληψίες επιφανειών καννάβων. Σε σχετική βιβλιογραφία το υπόβαθρο συναντάται και ως πλαίσιο.

1.7.3 Επιλογή δειγματοληπτικής μεθόδου

Το πρώτο στάδιο της δειγματοληπτικής διαδικασίας τελειώνει με την επιλογή του κατάλληλου δειγματοληπτικού υποβάθρου.

Ακολουθεί η επιλογή των παρατηρήσεων που θα αποτελέσουν το δείγμα. Οι επόμενες παράγραφοι αναφέρονται σε μεθόδους τυχαίας δειγματοληψίας (probability sampling) καθώς αυτές χρησιμοποιούνται περισσότερο, εξαιτίας και των βασικών πλεονεκτημάτων που παρουσιάζουν (τυχειότητα, μεγαλύτερη αντικειμενικότητα).

Η διαδικασία αυτή μπορεί να γίνει με τρεις βασικούς τρόπους: **Τυχαία (απλή), Συστηματική, Στρωματοποιημένη**. Υπάρχουν και διάφορες πιο σύνθετες μορφές, για πιο σύνθετα προβλήματα όπως η Ιεραρχική, η Πολυπαραγοντική, η Χωρο-Χρονική δειγματοληψία.

Στις επόμενες παραγράφους αρχικά λοιπόν αναλύονται οι βασικές αρχές τις κάθε μεθόδου και έπειτα επικεντρώνονται στην εφαρμογή τους σε χωρικές δειγματοληψίες.

1.8 Τυχαία δειγματοληψία

Μια διαδικασία δειγματοληψίας ονομάζεται απλή τυχαία εάν σε ένα δείγμα μεγέθους n , όλοι οι δυνατοί συνδυασμοί των n στοιχειωδών μονάδων που μπορούν να σχηματιστούν από τον πληθυσμό N στοιχειωδών μονάδων έχουν την ίδια πιθανότητα επιλογής.

Η τυχαία δειγματοληψία αποτελεί τη βάση κάθε στατιστικής επεξεργασίας στοιχείων και πληρεί δύο βασικές στατιστικές συνθήκες:

- Τα μέλη ενός πληθυσμού έχουν ίσες πιθανότητες να επιλεγούν για το δείγμα
- Η επιλογή ενός μέλους του πληθυσμού για το δείγμα με κανένα τρόπο δεν επηρεάζει την επιλογή ενός οποιουδήποτε άλλου.

Εξασφαλίζεται λοιπόν **ίση πιθανότητα** επιλογής για τις παρατηρήσεις και **ανεξαρτησία**. Στη περίπτωση που η χωρική διάσταση ενός πληθυσμού φαινομένων, αντικειμένων είναι υπαρκτή αλλά όχι και καθοριστική (περιοχή ως σύνολο), όπως προαναφέρθηκε, το δειγματοληπτικό υπόβαθρο μπορεί να είναι ένας κατάλογος. Η επιλογή ενός τυχαίου δείγματος από το υπόβαθρο αυτό γίνεται με τη χρήση πινάκων τυχαίων αριθμών.

Στη περίπτωση που ο αρχικός πληθυσμός είναι άπειρος, η λήψη ενός στοιχείου με επανάθεση, δηλαδή επανατοποθέτηση του επιλεγέντος στοιχείου κατά την προηγούμενη επιλογή, ή χωρίς επανάθεση, δεν αλλοιώνει την αρχική σύνθεση του πληθυσμού. Έτσι στη περίπτωση αυτή η λήψη του τυχαίου δείγματος συνεπάγεται ισονομία – ίση πιθανότητα επιλογής και ανεξαρτησία μεταξύ τους. Η συνάρτηση συχνότητας του δείγματος, δηλαδή η πιθανότητα να ληφθούν οι τιμές n_1, n_2, n_v θα είναι:

$$f(n_1, n_2, \dots, n_v) = \prod_{i=1}^v f_n(n_i)$$

Εάν ο αρχικός πληθυσμός είναι πεπερασμένος, τότε στη περίπτωση που η δειγματοληψία γίνεται με επανάθεση, οι τυχαίες μεταβλητές θα είναι ισόνομες και ανεξάρτητες. Εάν όμως η δειγματοληψία γίνεται χωρίς επανάθεση, οι τυχαίες μεταβλητές δεν είναι στατιστικά ανεξάρτητες.

Χρήση Πινάκων Τυχαίων Αριθμών

Για τη λήψη τυχαίας δειγματοληψίας χρησιμοποιούνται ευρύτατα στη πράξη οι πίνακες τυχαίων αριθμών, οι οποίοι κατασκευάζονται με βάση την αρχή της τυχαίας εκλογής με επανάθεση. Για παράδειγμα για τη λήψη ενός δείγματος, μεγέθους v , από ένα πεπερασμένο αριθμό N στοιχείων, με βάση την αρχή της τυχαίας εκλογής με επανάθεση, θα έπρεπε να τεθούν στη κάλπη N στοιχεία με συγκεκριμένη αρίθμηση και να γίνει εξαγωγή με επανάθεση των v στοιχείων του δείγματος. Για να αποφευχθεί η πρακτική επανάληψη του πειράματος χρησιμοποιούνται οι πίνακες των τυχαίων αριθμών, οι οποίοι περιέχουν, κατά σειρά, αριθμούς με ορισμένο πλήθος στοιχείων που επιλέχτηκαν με τυχαίο τρόπο. Στην περίπτωση μονοψήφιων αριθμών εξάγεται τυχαία ένα στοιχείο από τους αριθμούς 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 με επανάθεση και σημειώνεται ο αριθμός κατά σειρά εξαγωγής. Για διψήφιους εξάγονται δύο στοιχεία και σημειώνονται οι αντίστοιχοι αριθμοί. Για τετραψήφιους τυχαίους μπορούν να χρησιμοποιηθούν ανά δύο οι διψήφιοι

τυχαίοι που έχουν δημιουργηθεί κ.ο.κ. Για τη χρήση των πινάκων στη λήψη τυχαίας δειγματοληψίας ακολουθείται συνήθως η εξής διαδικασία:

Αριθμούνται όλα τα στοιχεία του πληθυσμού από τον οποίο θα ληφθεί το δείγμα και στη συνέχεια επιλέγεται μία τυχαία αφετηρία στο πίνακα και λαμβάνεται ένα πλήθος τυχαίων αριθμών ίσο προς το μέγεθος του δείγματος.

Ο τρόπος κατασκευής των πινάκων τυχαίων αριθμών ποικίλει και δεν υπάρχει δυνατότητα σύγκρισης μεταξύ τους. Ως βασικό κριτήριο καταλληλότητας των πινάκων είναι η κατά προσέγγιση ίση συχνότητα εμφάνισης των 10 βασικών αριθμών. Η κατασκευή τους είναι ιδιαίτερα εύκολη με τη χρήση Η/Υ.

Οι πίνακες τυχαίων αριθμών πληρούν τις δυο συνθήκες: ίση πιθανότητα – ανεξαρτησία. Βέβαια, η μέθοδος αυτή είναι επίπονη και για το λόγο αυτό συναντά πρακτικές δυσκολίες σε μεγάλες εφαρμογές (π.χ. μεγάλα δείγματα). Πίνακας τυχαίων αριθμών υπάρχει στο παράρτημα.

90057	12880	89522	81453	41294	14603
97431	01310	04978	25833	61253	48959
03346	61125	91434	84200	73664	45914
04812	76028	20044	78244	58015	05867
74532	60769	26928	29666	84318	21283
43016	56171	27506	72769	60939	50508
90227	88560	86908	32283	83958	92590
04809	23138	34852	45986	73823	08534
59332	70137	24535	17957	21012	51279
20195	14420	27684	35554	90414	80832
43144	85543	85639	91044	58951	85251
14952	50079	08851	56337	76769	83059
12700	60802	93248	12585	88835	97516
88126	63220	09217	89809	41827	63143
83873	30245	69153	36222	71188	50184

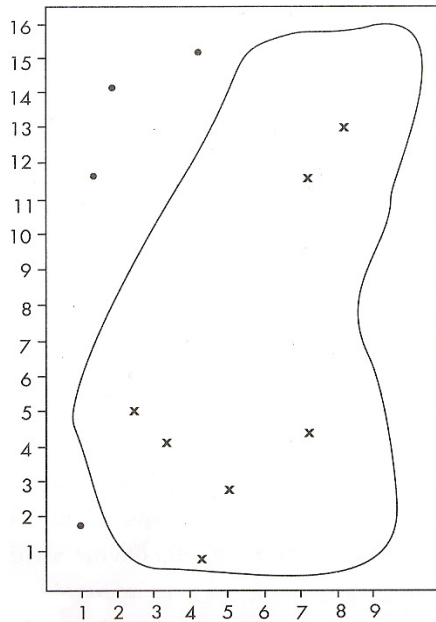
σχήμα 1.1: απόσπασμα πίνακα τυχαίων αριθμών

1.8.1 Σημειακή Δειγματοληψία

Στη περίπτωση της σημειακής δειγματοληψίας κάθε μέλος του πληθυσμού αντιπροσωπεύεται από ένα ζεύγος συντεταγμένων (x , y). Χρησιμοποιώντας ένα σύστημα συντεταγμένων πάνω στο δειγματοληπτικό υπόβαθρο, δημιουργείται ένα σύνολο (πληθυσμός) από σημεία που ορίζονται από τις συντεταγμένες τους.

Με τη χρήση πινάκων τυχαίων αριθμών γίνεται η επιλογή του τυχαίου δείγματος.

Όπως προαναφέρθηκε, ως αφετηρία για τη χρήση του πίνακα τυχαίων αριθμών επιλέγεται τυχαία ένας αριθμός και έπειτα με ένα προκαθορισμένο τρόπο (πχ. κάθε νιοστός αριθμός διαγωνίως) επιλέγονται οι υπόλοιποι έως τη συμπλήρωση του δείγματος. Σε περίπτωση που ο αριθμός που λαμβάνεται με τον τρόπο που αναφέρθηκε δεν βρίσκεται εντός του υποβάθρου ή επιλέγεται για δεύτερη φορά το σημείο αυτό δεν συμπεριλαμβάνεται στο δείγμα. Η σημειακή δειγματοληψία χρησιμοποιείται στις περισσότερες περιπτώσεις χαρτών που απεικονίζουν γεωλογικούς σχηματισμούς ή χρήσεις γης.

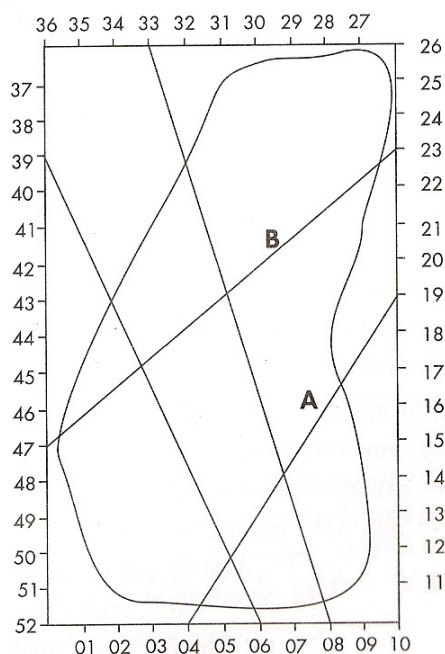


σχήμα 1.2: σημειακή τυχαία χωρική δειγματοληψία

1.8.2 Γραμμική Δειγματοληψία

Στη περίπτωση της γραμμικής δειγματοληψίας, η πληροφορία που αντλούμε λαμβάνεται μέσω ενός αριθμού γραμμών τυχαία κατανεμημένων στο χάρτη. Έτσι σε αντίθεση με τη σημειακή όπου ένα σημείο αντιστοιχεί σε μια χρήση γης, όταν ο χάρτης από όπου αντλούμε τη πληροφορία είναι χάρτης χρήσεων γης, στη γραμμική θα μιλάμε για ποσοστό της γραμμής που περνά μέσα από τις διάφορες χρήσεις. Οι γραμμές προέρχονται από τυχαία επιλογή.

Ένας τρόπος για τη δημιουργία τυχαίας γραμμικής δειγματοληψίας είναι να βαθμονομηθούν τα περιθώρια του υποβάθρου και στη συνέχεια με το πίνακα τυχαίων αριθμών να καθορίζεται η αρχή και το τέλος των γραμμών.

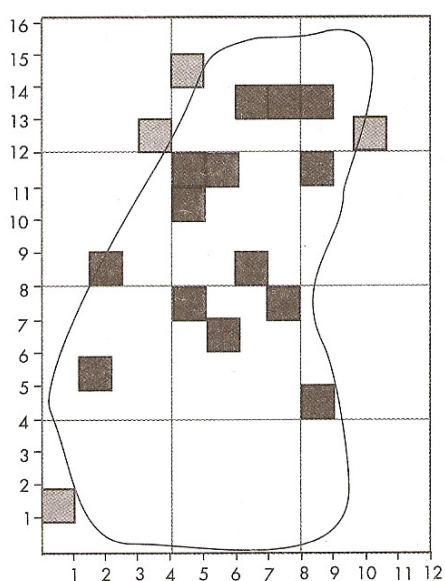


σχήμα 1.3: γραμμική τυχαία χωρική δειγματοληψία

1.8.3 Επιφανειακή Δειγματοληψία

Στη περίπτωση που το ενδιαφέρον εστιάζεται σε χωρικά φαινόμενα όπου η επιφάνεια αποτελεί συστατικό στοιχείο μπορεί να χρησιμοποιηθεί η επιφανειακή – κανναβική δειγματοληψία.

Στην επιφανειακή δειγματοληψία, το υπόβαθρο θεωρείται ότι αποτελείται από φατνία ενός καννάβου από όπου επιλέγεται τυχαία ένας μικρότερος αριθμός για να αποτελέσει το δείγμα. Αρχικά δημιουργείται ένα σύστημα συντεταγμένων που οριοθετεί το δειγματοληπτικό υπόβαθρο. Με το γνωστό τρόπο επιλέγονται από το πίνακα τυχαίων αριθμών ζεύγη τιμών που αντιστοιχούν στη κατω αριστερή γωνία του κάθε φατνίου. Καθοριστικής σημασίας για το αποτέλεσμα της δειγματοληψίας είναι το σχήμα και το μέγεθος του κάθε φατνίου. Συνήθως χρησιμοποιείται τετράγωνο μικρού μεγέθους.



σχήμα 1.4: επιφανειακή τυχαία χωρική δειγματοληψία

1.9 Συστηματική Δειγματοληψία

Δύο σπουδαία πλεονεκτήματα της τυχαίας δειγματοληψίας είναι η ανεξαρτησία και η ίση πιθανότητα επιλογής. Πρακτικά όμως η τυχαία δειγματοληψία παρουσιάζει το μειονέκτημα της ελλιπούς κάλυψης του δειγματοληπτικού υποβάθρου, με αποτέλεσμα κάποιες περιοχές του υποβάθρου να μην αντιπροσωπεύονται από σημεία δειγματοληψίας (στη σημειακή) και έτσι να υποβαθμίζεται η ύπαρξη τους.

Η συστηματική δειγματοληψία περιορίζει αυτό το μειονέκτημα της τυχαίας, αφού η επιλογή των επιμέρους παρατηρήσεων που αποτελούν το δείγμα γίνεται με ένα σταθερό και προκαθορισμένο (συστηματικό) τρόπο.

Στη συστηματική δειγματοληψία επιλέγεται ένα δείγμα μεγέθους n , από ένα πληθυσμό με N στοιχεία που είναι αριθμημένα από 1 έως N . Συνήθως, τίθεται η υπόθεση ότι $N = n * k$, όπου k φυσικός αριθμός. Με τη χρήση τυχαίων αριθμών επιλέγεται ένας αριθμός $1 \leq j \leq k$, και διαχωρίζονται ως δείγμα εκείνα τα στοιχεία του πληθυσμού που έχουν αύξοντες αριθμούς $j + h * k$, $h = 0, 1, 2, \dots, n - 1$

Η συστηματική δειγματοληψία βρίσκει εφαρμογή σε μη χωρικά αλλά και σε χωρικά υπόβαθρα. Όταν το υπόβαθρο είναι ένας κατάλογος (μη χωρικό), επιλέγεται ένα σταθερό

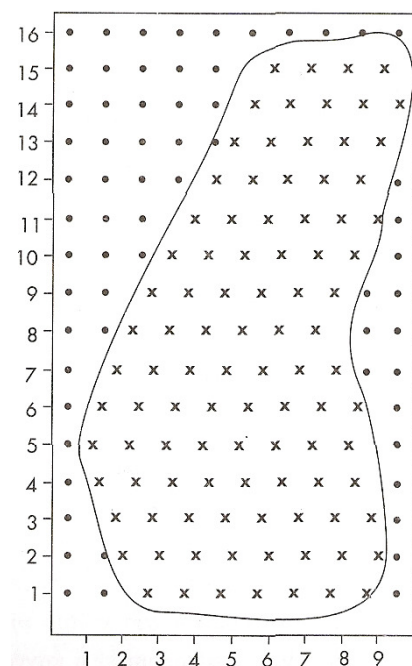
βήμα k , που ορίζεται ως το αντίστροφο της αναλογίας $\frac{\text{δειγμα}(n)}{\text{πληθυσμος}(N)}$. Το δείγμα

σχηματίζεται επιλέγοντας μια τυχαία αφετηρία, δηλαδή έναν αριθμό μεταξύ **1 και k** και στη συνέχεια επιλέγεται κάθε φορά ο **k** επόμενος αριθμός.

Για την περίπτωση των χωρικών υποβάθρων που συνήθως είναι κάποιος χάρτης ακολουθούν τα επόμενα:

1.9.1 Σημειακή Δειγματοληψία

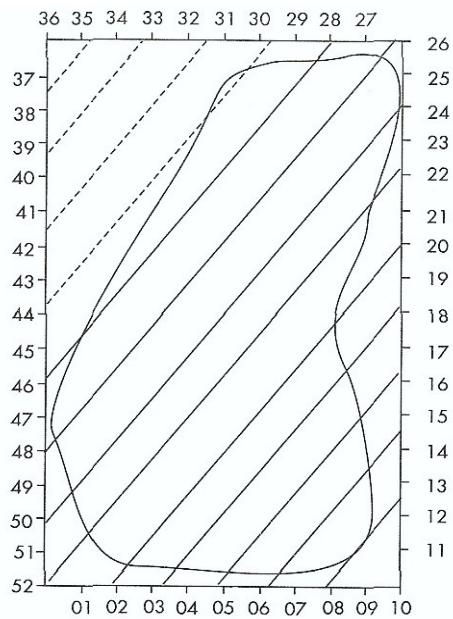
Η δειγματοληψία ξεκινά από ένα τυχαία επιλεγμένο σημείο (δύο αριθμοί από πίνακα τυχαίων αριθμών) ώστε να προσδιοριστούν οι συντεταγμένες του. Τα υπόλοιπα σημεία επιλέγονται με συστηματικό τρόπο, για παράδειγμα κάθε σημείο σε ίση απόσταση, κατά τους άξονες x και y , από το άλλο.



σχήμα 1.5: συστηματική σημειακή χωρική δειγματοληψία

1.9.2 Γραμμική δειγματοληψία

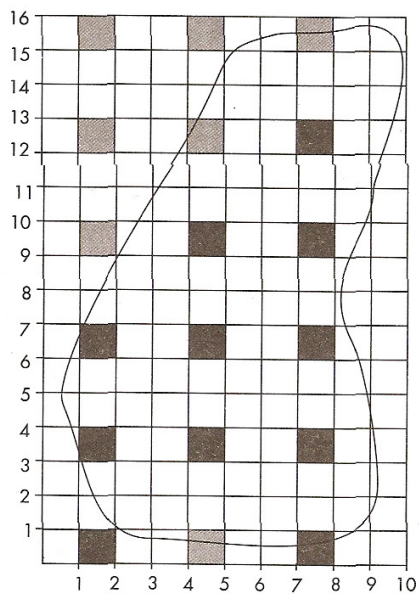
Ίδια μέθοδος ακολουθείται και στη γραμμική. Εδώ επιλέγεται τυχαία η πρώτη γραμμή και όλες οι υπόλοιπες γραμμές τοποθετούνται παράλληλες προς την αρχική σε μια συγκεκριμένη απόσταση (προκαθορισμένος τρόπος) μετρούμενη στον άξονα της αρχής της πρώτης γραμμής.



σχήμα 1.6: συστηματική γραμμική χωρική δειγματοληψία

1.9.3 Επιφανειακή (Κανναβική) Δειγματοληψία

Κατά αντιστοιχία με τα προηγούμενα στη περίπτωση της επιφανειακής συστηματικής, αρχικά επιλέγεται η πρώτη επιφάνεια του δείγματος τυχαία (όπως στην επιφανειακή τυχαία δειγματοληψία) και έπειτα τα επόμενα μέλη του δείγματος επιλέγονται με τέτοιο τρόπο ώστε η κάτω αριστερή γωνία κάθε φατνίου να βρίσκεται σε μια σταθερή και προκαθορισμένη απόσταση, κατά τους άξονες x και y , από το προηγούμενο φατνίο.



σχήμα 1.7: συστηματική επιφανειακή χωρική δειγματοληψία

Η συστηματική δειγματοληψία παρέχει το σημαντικό πλεονέκτημα της ομοιόμορφης κάλυψης του δειγματοληπτικού υποβάθρου. Υπάρχουν βέβαια και δύο σημαντικά μειονεκτήματα.

Αν ο πληθυσμός περικλείει έναν περιοδικό τύπο διακύμανσης και εάν το διάστημα μεταξύ διαδοχικών μονάδων δειγματοληψίας συμπίπτει με το μήκος κύματος της περιοδικότητας, μπορεί να πάρουμε ένα δείγμα πάρα πολύ μεροληπτικό. Για παράδειγμα ένα συστηματικό δείγμα των κατοικιών μιας πόλης θα μπορούσε να περιέχει πάρα πολλές ή πολύ λίγες γωνιακές κατοικίες.

Το άλλο βασικό μειονέκτημα είναι ότι το αρχικό σημείο καθορίζει και την επιλογή των υπολοίπων σημείων. Έτσι μετά τη τυχαία επιλογή του πρώτου σημείου, όλα τα σημεία του υποβάθρου δεν έχουν την ίδια πιθανότητα επιλογής, αφού τα υπόλοιπα επιλέγονται με συστηματικό τρόπο.

Ωστόσο η συστηματική δειγματοληψία συνήθως είναι μέρος ενός περισσότερο σύνθετου σχεδίου στο οποίο είναι δυνατό να λάβουμε αμερόληπτες εκτιμήσεις. Στα σχέδια αυτά η τυχαία διαδικασία λαμβάνει μέρος όχι μόνο για τον προσδιορισμό ενός αρχικού σημείου (στα χωρικά υπόβαθρα) ή ενός μέλους του καταλόγου (στα μη χωρικά) αλλά και καθ όλη την εφαρμογή της.

1.10 Στρωματοποιημένη Δειγματοληψία (Απλή Τυχαία)

Στη στρωματοποιημένη δειγματοληψία ο πληθυσμός N διαιρείται σε k υποπληθυσμούς (στρώματα): $N_1, N_2, N_3, \dots, N_k$ έτσι ώστε να μην υπάρχει επικάλυψη μεταξύ των υποπληθυσμών (στρωμάτων) και το σύνολο των επιμέρους στρωμάτων να συνιστούν τον αρχικό πληθυσμό N . [$N_1 + N_2 + \dots + N_k = N$]

Η διαδικασία κατά την οποία γίνεται λήψη απλού τυχαίου δείγματος από κάθε στρώμα είναι γνωστή σαν στρωματοποιημένη δειγματοληψία (ή κατά στρώματα τυχαία).

Η στρωματοποιημένη δειγματοληψία είναι ιδιαίτερα σημαντική όταν ο πληθυσμός αποτελείται από διακεκριμένες υποομάδες που χαρακτηρίζονται από διαφορετικό μέγεθος ή χαρακτηριστικά.

Διάφορες καταστάσεις προκρίνουν τη στρωματοποιημένη δειγματοληψία, για παράδειγμα για τον καθορισμό της έκτασης που καταλαμβάνουν τα εμπορικά καταστήματα θα πρέπει να δημιουργηθούν περιφέρειες γύρω από το κέντρο της πόλης και στη συνέχεια να ληφθούν δείγματα και αυτό γιατί η έκταση που καταλαμβάνουν ελαττώνεται με την απομάκρυνση από το κέντρο και έτσι ανάλογα πρέπει να ληφθούν και τα δείγματα – από κάθε στρώμα (περιφέρεια) - για να απεικονιστεί η πραγματικότητα.

Επίσης η σύγκριση κάποιου (χωρικού) χαρακτηριστικού δύο ή περισσότερων περιοχών, όπως και η σύγκριση δύο μεταβλητών όταν δεν διαφοροποιούνται ομοιόμορφα στο χώρο, αποτελούν παραδείγματα στα οποία η στρωματοποίηση του δείγματος είναι μια σωστή προσέγγιση.

Χαρακτηριστικό πλεονέκτημα της στρωματοποίησης είναι η αντιπροσώπευση των διαφόρων υποομάδων του πληθυσμού στο δείγμα. Παράλληλα, αυξάνεται και η ποιότητα των εκτιμήσεων για τον πληθυσμό.

Στη χωρική δειγματοληψία που επιλέγεται η στρωματοποίηση του υποβάθρου, υπάρχουν δύο προσεγγίσεις:

- φυσική στρωματοποίηση
- κανναβική στρωματοποίηση

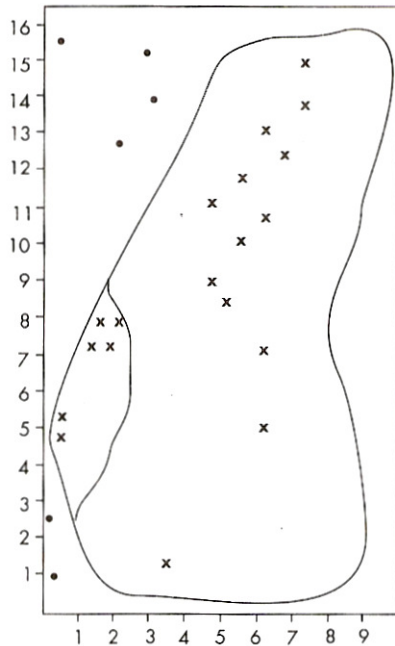
Στη **φυσική στρωματοποίηση**, μπορούν να διακριθούν υποπεριοχές γνωστής έκτασης στο δειγματοληπτικό υπόβαθρο. Οι υποπεριοχές αυτές θα αποτελέσουν τα στρώματα, στα οποία θα γίνει ανεξάρτητη λήψη δειγμάτων.

Συνήθως, το μέγεθος του δείγματος σε κάθε στρώμα – υποπεριοχή είναι ανάλογο της επιφάνειας της, που είναι γνωστή.

Στη **κανναβική στρωματοποίηση**, η περιοχή μελέτης επικαλύπτεται από ένα κάνναβο με φατνία ίσου μεγέθους και για κάθε υποπεριοχή – στρώμα υπολογίζεται το μέγεθος του δείγματος το οποίο καθορίζει τον αριθμό των σημείων σε κάθε φατνίο. Η επιλογή των σημείων σε κάθε φατνίο γίνεται με το γνωστό τρόπο τυχαίας δειγματοληψίας.

1.10.1 Στρωματοποιημένη τυχαία δειγματοληψία

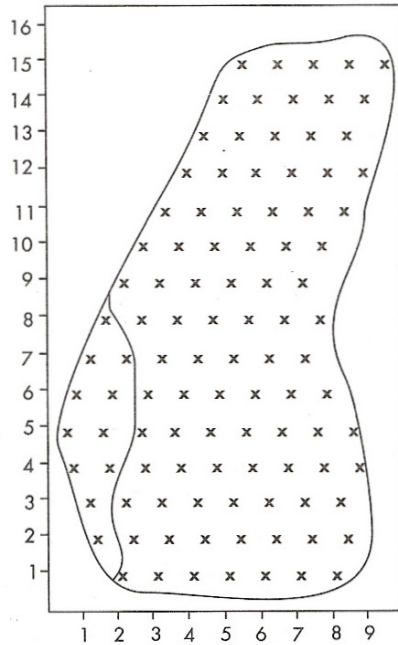
Αρχικά επιλέγεται το υπόβαθρο και διαιρείται σε υποπεριοχές που θα αποτελούν τα στρώματα της δειγματοληψίας. Σε καθεμιά από αυτές τις υποπεριοχές γίνεται τυχαία δειγματοληψία, με αριθμό σημείων – εάν πρόκειται για σημειακή – που έχει ήδη καθοριστεί (πχ. ανάλογα με τη επιφάνεια κάθε στρώματος).



σχήμα 1.8: στρωματοποιημένη τυχαία δειγματοληψία

1.10.2 Στρωματοποιημένη συστηματική δειγματοληψία

Στη περίπτωση αυτή η διαφορά με τη τυχαία στρωματοποιημένη δειγματοληψία έγκειται στον τρόπο με τον οποίο γίνεται η επιλογή του δείγματος στις υποπεριοχές – στρώματα. Η με συστηματικό τρόπο επιλογή του δείγματος στις υποπεριοχές του υποβάθρου προσφέρει τα πλεονεκτήματα της συστηματικής δειγματοληψίας που έχουν αναφερθεί, για παράδειγμα η πλήρης κάλυψη του δειγματικού υποβάθρου.



σχήμα 1.9: στρωματοποιημένη συστηματική δειγματοληψία

1.10.3 Στρωματοποιημένη συστηματική μη γραμμική δειγματοληψία

Το πιο σύνθετο αυτό είδος δειγματοληψίας συνδυάζει τα πλεονεκτήματα της τυχαίας επιλογής, της στρωματοποίησης και του συστηματικού δείγματος. Η σύνθετη διαδικασία εφαρμογής της περιλαμβάνει αρχικά τη διαίρεση του υποβάθρου με βάση ένα σύστημα συντεταγμένων σε κάναβο με φατνία ίσου μεγέθους. Η επιλογή του πρώτου σημείου γίνεται στο πρώτο γωνιακό φατνίο με τυχαίο τρόπο. Η επιλογή των επόμενων σημείων – ένα σημείο σε κάθε φατνίο - καθορίζεται από τις συντεταγμένες του πρώτου (συστηματικότητα). Συγκεκριμένα, τα σημεία των φατνίων της πρώτης γραμμής θα έχουν ίδια τετμημένη με αυτή του πρώτου και τεταγμένες που στο καθένα επιλέγονται τυχαία (από πίνακα τυχαίων αριθμών). Παράλληλα, όλα τα σημεία της πρώτης στήλης φατνίων θα έχουν την ίδια τεταγμένη με αυτή του πρώτου σημείου και τετμημένες που επιλέγονται με τυχαίο τρόπο στο καθένα. Η επιλογή των επόμενων σημείων έχει πλέον καθοριστεί.

1.11 Εφαρμογή της Κανονικής Κατανομής

Σε πολλές περιπτώσεις δειγματοληψιών, τόσο ο αρχικός πληθυσμός, όσο και τα επιλεγόμενα δείγματα, είναι αρκετά μεγάλα. Το γεγονός αυτό θεμελιώνει ένα πολύ ουσιαστικό στοιχείο της Θεωρίας των Πιθανοτήτων που αποτελεί το Κεντρικό Οριακό Θεώρημα και που οδηγεί στην εφαρμογή της κανονικής κατανομής για ένα άθροισμα ανεξάρτητων τυχαίων μεταβλητών.

Η κανονική κατανομή είναι η σημαντικότερη κατανομή πιθανοτήτων με τις περισσότερες εφαρμογές. Μελετήθηκε αρχικά από τον De Moivre (1667 – 1754) και από τον Laplace (1749 – 1827), οι οποίοι απέδειξαν ότι η διωνυμική κατανομή $B(n,p)$ προσεγγίζεται, για μεγάλο n , από την κανονική κατανομή. Το 1809 ο Gauss (1777 – 1855), μελετώντας τη θεωρία των τυχαίων σφαλμάτων, παρατήρησε ότι οι κατανομές των σφαλμάτων των μετρήσεων μπορούσαν ικανοποιητικά να προσεγγιστούν από μία συνεχή καμπύλη, η οποία αναφερόταν ως κανονική καμπύλη των σφαλμάτων και αποδίδονταν στους νόμους της τύχης.

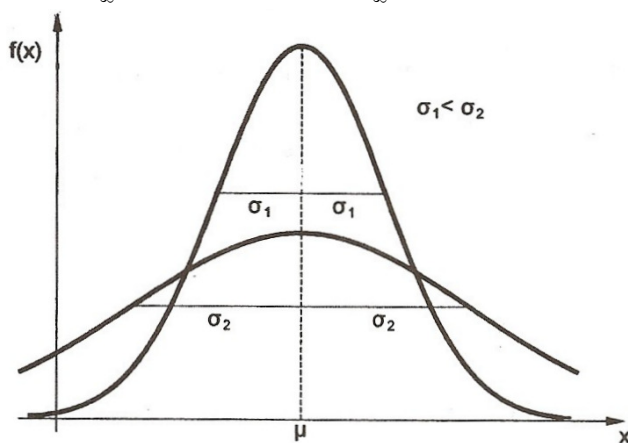
Η συνάρτηση πυκνοτήτων της κανονικής κατανομής είναι:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{για } -\infty < x < +\infty$$

Οι παράμετροι μ , σ είναι η ακριβής τιμή της τυχαίας μεταβλητής και το τυπικό σφάλμα των μετρήσεων.

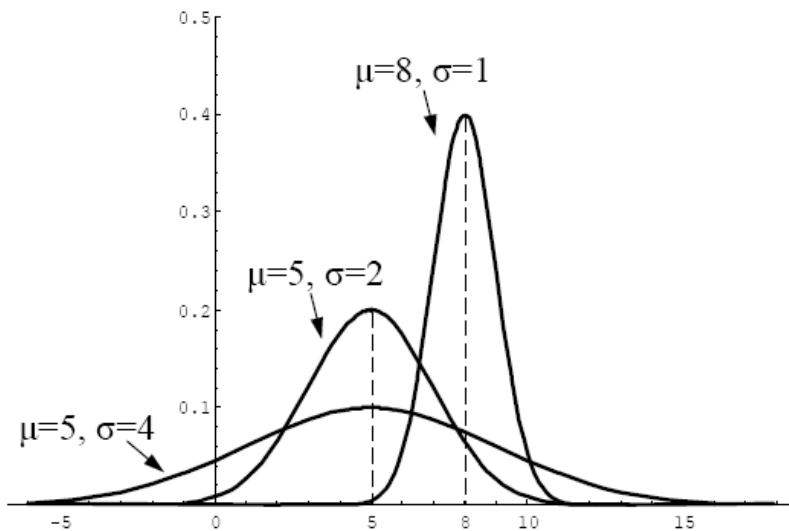
Η συνάρτηση κατανομής της κανονικής κατανομής είναι η:

$$F(x) = \int_{-\infty}^x f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$



σχήμα 1.10: γραφικές παραστάσεις συναρτήσεων πυκνοτήτων κανονικής κατανομής ($\sigma_1 < \sigma_2$).

Η συνάρτηση $f(x)$ είναι συμμετρική ως προς την ακριβή τιμή μ , ενώ τα σημεία καμπής της βρίσκονται στις θέσεις $x = \mu - \sigma$ και $x = \mu + \sigma$.



σχήμα 1.11: γράφημα της συνάρτησης πυκνότητας πιθανότητας για διάφορα (μ, σ) .

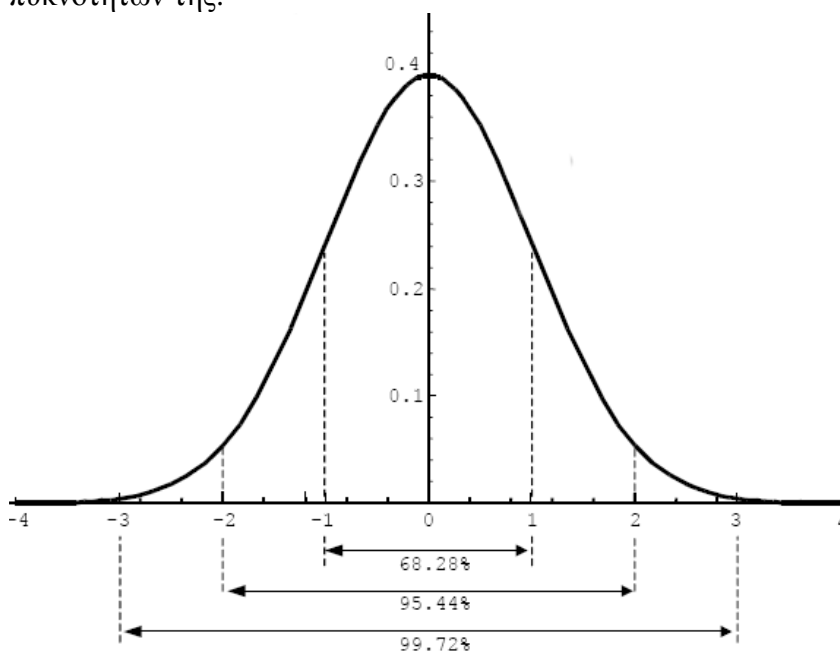
Τυποποιημένη κανονική κατανομή

Αν αντί της τυχαίας μεταβλητής x θεωρήσουμε την $z = \frac{x - \mu}{\sigma}$ καταλήγουμε σε ακριβή τιμή $\mu = 0$ και μεταβλητότητα $\sigma = 1$ έτσι:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz = F(x)$$

Η τυχαία μεταβλητή z ονομάζεται τυποποιημένη μεταβλητή και η συνάρτηση πυκνοτήτων της:



Σχήμα 1.12: συνάρτηση πυκνότητας πιθανότητας της τυπικής κανονικής κατανομής.

Όπως φαίνεται η $f(z)$ προκύπτει από την $f(x)$ αν τεθεί $\mu = 0$ και $\sigma = 1$.

Έτσι οι $f(z)$ και $F(z)$ πινακοποιούνται για $z \geq 0$ και από αυτές υπολογίζονται οι πιθανότητες που ενδιαφέρουν.

Είναι προφανές ότι: $f(-z) = f(z)$ και $F(-z) = 1 - F(z)$

ενώ $P(-z_i \leq z \leq z_i) = 2F(z_i) - 1$

1.11.1 Μέτρα ακρίβειας – Διαστήματα εμπιστοσύνης

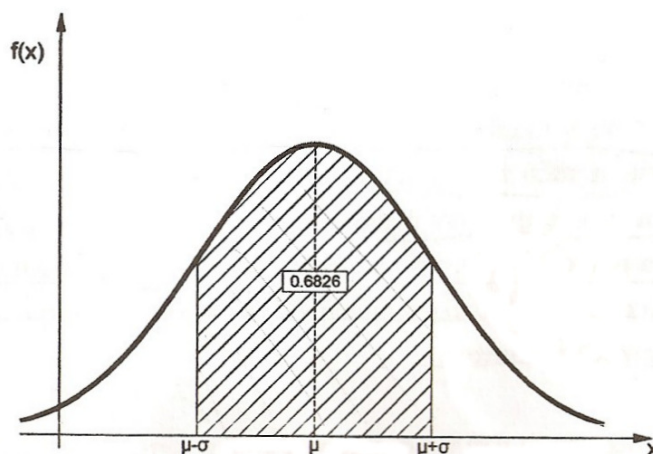
Ένα μέτρο ακρίβειας αποτελεί το τυπικό σφάλμα σ , καθώς όσο μεγαλώνει το σ τόσο πιο πεπλατυσμένη γίνεται η συνάρτηση πυκνοτήτων $f(x)$, κάτι που σημαίνει ότι η ακρίβεια είναι αντίστροφη του τυπικού σφάλματος.

Έτσι δικαιολογείται και το όνομα τυπικό σφάλμα σαν ένα μέγεθος που χαρακτηρίζει (τυποποιεί) μια σειρά μετρήσεων για παράδειγμα.

Θεωρώντας την τυποποιημένη μεταβλητή z προκύπτει:

$$P(\mu - \sigma < x < \mu + \sigma) = P(-1 < z < 1) = 0.6826$$

Από όπου συμπεραίνεται ότι η διαφορά μιας εκτίμησης x της μεταβλητής από την ακριβή τιμή της θα κυμαίνεται μεταξύ $-\sigma$ και $+\sigma$ (αν x είναι η μέση τιμή τότε σ θα είναι το τυπικό σφάλμα της μέσης τιμής) με πιθανότητα περίπου 68%. Κάτι που σημαίνει πως αν η εκτίμηση που έχουμε κάνει είναι x , η ακριβής τιμή θα βρίσκεται το διάστημα $x - \sigma$ έως $x + \sigma$ με πιθανότητα 68%.



Για μέτρα ακρίβειας χρησιμοποιούνται και πολλαπλάσια του τυπικού σφάλματος, όπως το 2σ ή 3σ , για τα οποία και βρίσκονται οι αντίστοιχες πιθανότητες.

$$P(\mu - 2\sigma < x < \mu + 2\sigma) = P(-2 < z < 2) = 95.4\%$$

$$P(\mu - 3\sigma < x < \mu + 3\sigma) = P(-3 < z < 3) = 99.7\%$$

Συχνά παρουσιάζεται το πρόβλημα να βρεθεί το διάστημα μέσα στο οποίο αναμένεται να βρίσκεται η διαφορά $x - \mu$ με ορισμένη πιθανότητα P .

Η πιθανότητα ονομάζεται **επίπεδο εμπιστοσύνης** και το διάστημα που συνδέεται με αυτή **διάστημα εμπιστοσύνης**.

z_p	$P = P(x - z\sigma < \mu < x + z\sigma) = 2F(z) - 1$
0.6745	0.500
1.000	0.683
1.645	0.900
1.960	0.950
2.000	0.954
2.576	0.990
3.000	0.997

σχήμα 1.13: τιμές Z για διάφορα διαστήματα εμπιστοσύνης

Έτσι π.χ. για επίπεδο εμπιστοσύνης $p = 99\%$, έχουμε $Z_p = 2.576$ που σημαίνει διάστημα εμπιστοσύνης για τη διαφορά $x - \mu$ από -2.576σ έως 2.576σ ή

$$P(x - 2.576\sigma < \mu < x + 2.576\sigma) = 99\%$$

Το επίπεδο εμπιστοσύνης γενικά συμβολίζεται και με $p = 1 - \alpha$ όπου α η πιθανότητα να βρίσκεται η $x - \mu$ έξω από το διάστημα εμπιστοσύνης.

Ο όρος παράμετρος πληθυσμού (population parameter) χρησιμοποιείται για να δηλωθεί το χαρακτηριστικό ενός πληθυσμού, ενώ ο όρος εκτιμητής δείγματος (sample statistic) για να δηλωθεί το χαρακτηριστικό ενός δείγματος. Η δειγματοληψία στηρίζεται στον προσδιορισμό εκτιμητών του δείγματος που στη περίπτωση που το δείγμα αυτό είναι αντιπροσωπευτικό και αμερόληπτο θα αποτελεί εκτίμηση για την αντίστοιχη παράμετρο του πληθυσμού.

Ακριβής εκτίμηση των παραμέτρων του πληθυσμού (μ , σ) (ακριβής τιμή και τυπική απόκλιση του πληθυσμού) από τα M και s (μέση τιμή και τυπική απόκλιση του δείγματος), δηλαδή τους εκτιμητές του δείγματος είναι αδύνατη, όπως αντίστοιχα δύο δείγματα να έχουν ίδιες τιμές για τα M και s .

Η κατανομή των εκτιμητών των δειγμάτων (πχ η μέση τιμή των δειγμάτων που λαμβάνονται από έναν πληθυσμό) είναι η δειγματοληπτική κατανομή τους. Αν τα δείγματα είναι αντιπροσωπευτικά και αμερόληπτα αναμένεται ο κάθε εκτιμητής τους να κατανέμεται συμμετρικά γύρω από την αντίστοιχη παράμετρο του πληθυσμού.

Ένα παράδειγμα: Έστω ότι από τον έλεγχο ενός τυχαίου δείγματος εκτιμάται ότι το 10% των χαρτών μιας συγκεκριμένης έκδοσης είναι ελαττωματικοί. Απέναντι σε αυτό το νούμερο, απόλυτα λογικές και αναμενόμενες είναι ερωτήσεις όπως: «μπορούμε να ισχυριστούμε με βεβαιότητα ότι το πραγματικό ποσοστό ελαττωματικών χαρτών κυμαίνεται μεταξύ 5-15%;» ή «πόσο πιθανόν είναι το ποσοστό των ελαττωματικών χαρτών να είναι μεταξύ 9-11%;»

1.11.2 Εκτιμήσεις διαστήματος εμπιστοσύνης για μετρήσεις

Η διαδικασία εκτίμησης βασίζεται σε ένα θεμελιώδες θεώρημα που ονομάζεται **θεώρημα του κεντρικού ορίου** (central limit theorem), το οποίο λέει ότι εάν ένα τυχαίο δείγμα n παρατηρήσεων ληφθεί από έναν πληθυσμό με ακριβή τιμή μ και μεταβλητότητα σ^2 , τότε, εφόσον οι παρατηρήσεις n είναι πολλές, η δειγματοληπτική κατανομή των μέσων τιμών θα προσεγγίζει την κανονική κατανομή.

Έτσι το τυπικό σφάλμα (όπως ονομάζεται η τυπική απόκλιση των δειγματοληπτικών κατανομών) θα δίνεται από:

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

όπου: σ_M : το τυπικό σφάλμα της δειγματοληπτικής κατανομής των M

σ : η τυπική απόκλιση του πληθυσμού

n : το μέγεθος του δείγματος

Στην τυποποιημένη της μορφή:

$$Z_M = \frac{M - \mu}{\sigma_M} = \frac{M - \mu}{\sigma/\sqrt{n}}$$

Η Z_M για επίπεδο εμπιστοσύνης $(100 - \alpha) \%$:

$$P(-Z_{\alpha/2} \leq Z_M \leq Z_{\alpha/2}) = 100 - \alpha$$

$$P(M - Z_{\alpha/2}\sigma_M \leq \mu \leq M + Z_{\alpha/2}\sigma_M) = 100 - \alpha$$

Έτσι το διάστημα μέσα στο οποίο βρίσκεται η πραγματική τιμή του πληθυσμού, είμαστε $(100-\alpha)\%$ σίγουροι ότι είναι:

$$M \pm Z_{\alpha/2}\sigma_M = M \pm Z_{\alpha/2}(\sigma/\sqrt{n})$$

όπου: $Z_{\alpha/2}$: η τιμή από τους πίνακες κανονικής κατανομής που αντιστοιχεί στην αθροιστική συχνότητα $100 - (\alpha/2)$.

Καθότι η τυπική απόκλιση του πληθυσμού δεν είναι γνωστή (σ) χρησιμοποιείται ένας εκτιμητής της τυπικής απόκλισης του πληθυσμού, η τυπική απόκλιση του δείγματος (s), έτσι έχουμε:

$$M \pm Z_{\alpha/2}(s/\sqrt{n}).$$

1.11.3 Εκτιμήσεις διαστήματος εμπιστοσύνης για απαριθμήσεις

Όταν οι παρατηρήσεις δεν αφορούν μετρήσεις (πχ μέσο μέγεθος επιφάνειας κατοικιών) αλλά απαριθμήσεις (πχ αριθμός κτιρίων σε χρήση ‘‘αμιγής κατοικία’’), τότε αυτό που ενδιαφέρει είναι η αναλογία των διαφόρων κατηγοριών στο δείγμα, έτσι ώστε να βγάλουμε ασφαλή συμπεράσματα για τις αναλογίες - χρήσεων των κτιρίων – ολόκληρου του πληθυσμού.

Η δειγματοληπτική κατανομή που χρησιμοποιείται είναι η διωνυμική, η οποία σχετίζεται με την επανάληψη τυχαίων γεγονότων που μπορούν να λάβουν δύο μόνο μορφές. (κτίρια σε βιομηχανική χρήση και κτίρια σε μη βιομηχανική χρήση).

Το διάστημα εμπιστοσύνης θα δίνεται από τη σχέση:

$$np \pm Z_{\alpha/2} \sqrt{npq}$$

Επειδή το ενδιαφέρον εστιάζεται στις αναλογίες των μεγεθών αυτών το διάστημα εμπιστοσύνης δίνεται σε επί τοις εκατό αναλογίες:

$$np \pm Z_{\alpha/2} \sqrt{100p \cdot 100q/n}$$

1.12 Μέγεθος του δείγματος και εκτίμηση του

Η απαρίθμηση και η απογραφή του πληθυσμού προσφέρει προφανώς τη μέγιστη ακρίβεια για όποιες παραμέτρους και χαρακτηριστικά του μελετώνται (φυσικά αυτό υπό τη προϋπόθεση ότι είναι δυνατό να διενεργηθεί απογραφή από εξειδικευμένο και έμπειρο προσωπικό). Όπως έχει αναφερθεί σε πολλές περιπτώσεις είναι αδύνατο να πραγματοποιηθεί απαρίθμηση – απογραφή του πληθυσμού καθώς είναι και αντιοικονομικό (σε χρόνο και χρήμα). Αυτό επιβάλλει την επιλογή ενός μικρού αριθμού των μελών του πληθυσμού για τη μελέτη των διαφόρων χαρακτηριστικών του (επιλογή δείγματος). Το δείγμα αυτό θα πρέπει να είναι αντιπροσωπευτικό του πληθυσμού, ώστε οι εκτιμήσεις που θα γίνονται για το σύνολο του πληθυσμού να είναι ασφαλείς. Καθώς όμως το δείγμα ελαττώνεται, το περιθώριο λάθους στις εκτιμήσεις για τον πληθυσμό μεγαλώνει. Αυτό που πρέπει λοιπόν να καθοριστεί είναι το **ελάχιστο μέγεθος του δείγματος** που ικανοποιεί ορισμένες απαιτήσεις ακρίβειας.

Πριν τη διενέργεια της δειγματοληψίας, λαμβάνεται συνήθως ένα δοκιμαστικό δείγμα λίγων παρατηρήσεων, με την ίδια μέθοδο (πχ απλό τυχαίο δείγμα 30 παρατηρήσεων), Το δοκιμαστικό μπορεί να αποτελεί και μέρος του οριστικού δείγματος. Η δοκιμαστική είναι μια δειγματοληψία μειωμένης ακρίβειας που δίνει μια εικόνα για τα χαρακτηριστικά του πληθυσμού που μελετάται και τα αποτελέσματα της χρησιμοποιούνται για τον υπολογισμό του οριστικού δείγματος.

Ο τρόπος εύρεσης του ελάχιστου μεγέθους δείγματος διαχωρίζεται όπως προαναφέρθηκε στα διαστήματα εμπιστοσύνης ανάλογα με το αν η δειγματοληψία αναφέρεται σε **μετρήσεις** ή σε **απαριθμήσεις**.

Στη περίπτωση που η δειγματοληψία αφορά **μετρήσεις** (πχ. εκτίμηση μέσου μεγέθους ιδιοκτησιών), το απαιτούμενο μέγεθος δείγματος (ή ελάχιστο μέγεθος δείγματος) εξαρτάται από το **περιθώριο λάθους θ** που θα θέσουμε καθώς και το **διάστημα εμπιστοσύνης** που θα επιλέξουμε. Η δοκιμαστική δειγματοληψία που έχει προηγηθεί έχει δώσει τιμή για την τυπική απόκλιση. Σύμφωνα λοιπόν με τη σχετική βιβλιογραφία το απαιτούμενο μέγεθος δείγματος για μετρήσεις δίνεται από τον τύπο:

$$n = (Z_{\alpha/2} \cdot s / \theta)^2$$

Όπου:

- n:** το απαιτούμενο μέγεθος του δείγματος
- $Z_{\alpha/2}$:** η τιμή του Z για διάστημα εμπιστοσύνης $(100 - \alpha)\%$
- S:** η τυπική απόκλιση του δείγματος
- θ :** το περιθώριο λάθους

Στη περίπτωση που η δειγματοληψία αφορά **απαριθμήσεις** – αναλογίες (εκτίμηση ποσοστών χρήσεων γης μιας περιοχής), το απαιτούμενο μέγεθος δείγματος δίνεται από το τύπο:

$$n = 100p \cdot 100q \cdot (Z_{\alpha/2} / \theta)^2$$

Όπου:

p και q: οι αναλογίες του δείγματος που ανήκουν και δεν ανήκουν στη συγκεκριμένη κατηγορία.

1.13 Έλεγχος Υποθέσεων – Test X^2 (καλής προσαρμογής)

Από τη μελέτη ενός δείγματος μπορούμε να εξάγουμε κάποια συμπεράσματα σχετικά με τα χαρακτηριστικά του που τα ανάγουμε στο σύνολο του πληθυσμού είτε με τον υπολογισμό των σημειακών εκτιμητριών είτε με τη χρήση διαστημάτων εμπιστοσύνης. Εναλλακτικά πληροφορίες που προέρχονται από ένα δείγμα μπορούν να χρησιμοποιηθούν ώστε να ελεγχθεί ως προς την αξιοπιστία της μια υπόθεση που ο ερευνητής έχει κάνει για τον πληθυσμό που εξετάζει.

Έτσι λοιπόν γίνεται μια υπόθεση αναφορικά με έναν πληθυσμό και τα συμπεράσματα για την ορθότητα αυτής της υπόθεσης βασίζονται στην επεξεργασία πληροφοριών που προέρχονται από το δείγμα.

Η διαδικασία που ακολουθείται είναι η εξής:

Έστω ότι η πληθυσμιακή παράμετρος που μας ενδιαφέρει συμβολίζεται με θ και κάποια υπόθεση διατυπώνεται για αυτήν τη παράμετρο. Θεωρούμε ότι η υπόθεση αυτή είναι ορθή εκτός και αν υπάρχουν ενδείξεις για το αντίθετο. Λαμβάνοντας στοιχεία από τυχαίο δείγμα θέτουμε σε αμφισβήτηση τη **μηδενική υπόθεση (H_0)**, που σε περίπτωση που αποδειχθεί αναληθής πρέπει να απορριφθεί και να αντικατασταθεί από μια άλλη υπόθεση (**H_1**), η οποία έχει σαφώς διατυπωθεί αρχικά.

Η παραπάνω διαδικασία στη στατιστική ονομάζεται **έλεγχος υποθέσεων**.

Εκτός των παραμετρικών υποθέσεων, ιδιαίτερα χρησιμοποιούνται οι έλεγχοι καλής προσαρμογής (goodness of fit test), μια πρώτη προσέγγιση έγινε από τον Karl Pearson στις αρχές του 20^{ου} αιώνα.

Το κριτήριο που αφορά τον έλεγχο των υποθέσεων:

$$H_0 : p_i = p_{oi} \quad (i = 1, \dots, k) \text{ μηδενική υπόθεση}$$

$$H_1 : p_i \neq p_{oi} \quad \text{για ένα τουλάχιστον } i \text{ εναλλακτική υπόθεση – απόρριψη μηδενικής}$$

όπου p_i ($i = 1, \dots, k$) οι πιθανότητες k αμοιβαίως αποκλειομένων ενδεχομένων A_i

Σε n ανεξάρτητες επαναλήψεις εάν N_i ($i = 1, \dots, k$) είναι οι αριθμοί εμφανίσεων των ενδεχομένων A_i $\left(\sum_1^k N_i = n \right)$, τότε η κατανομή της πολυδιάστατης τυχαίας μεταβλητής $N = (N_1, \dots, N_k)$ είναι η $(k - 1)$ - διάστατη Πολυωνυμική κατανομή.

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (\text{test statistic})$$

όπου: O_i : Η παρατηρούμενη συχνότητα της i κατηγορίας

E_i : Η αναμενόμενη / πραγματική συχνότητα της i κατηγορίας

k : Ο αριθμός των κατηγοριών

Περιοχή απόρριψης: $\chi^2 > \chi_\alpha^2$ όπου χ_α^2 έχει $(k - 1)$ βαθμούς ελευθερίας

Απαραίτητη προϋπόθεση για έχει ισχύ η προσέγγιση του test: η αναμενόμενη τιμή για κάθε κατηγορία να είναι μεγαλύτερη από 5. [$E(n_i) \geq 5$]

Εάν η αναμενόμενη συχνότητα μιας κατηγορίας είναι μικρότερη από 5 συγχωνεύεται με άλλη κατηγορία.

Παράδειγμα:

Συγκρίνουμε το μερίδιο αγοράς πριν και μετά την διαφημιστική καμπάνια, για να δούμε αν υπάρχει διαφορά.

Μερίδια αγοράς πριν από διαφημιστική εκστρατεία

ΕΤΑΙΡΙΕΣ	Μερίδιο Αγοράς Πριν %
A	45
B	40
Λοιπές	15

200 πελάτες καταμετρήθηκαν μετά την εκστρατεία, τα αποτελέσματα είναι:

ΕΤΑΙΡΙΕΣ	Αριθμός πελατών
A	102
B	82
Λοιπές	16

Εξετάζουμε αν η εκστρατεία ήταν αποτελεσματική σε συγκεκριμένο επίπεδο σημαντικότητας α .

- Η μηδενική υπόθεση είναι (H_0): $P(\text{εταιρίαA}) = 0.45$, $P(\text{εταιρίαB}) = 0.40$ $P(\text{λοιπές}) = 0.15$
- Η εναλλακτική υπόθεση (H_a): τουλάχιστον μία P_i διαφορετική.

Το στατιστικό τεστ που χρησιμοποιούμε είναι το test χ^2

Η περιοχή απόρριψης $\chi^2 > \chi^2_\alpha$

Τελικά βρίσκουμε $\chi^2 = 8.18$ και για επίπεδο σημαντικότητας $\alpha = 0.5$ και βαθμούς ελευθερίας $d = 3 - 1 = 2$, από πίνακα $\chi^2_\alpha = 5.99$

Έτσι $\chi^2 > \chi^2_\alpha$ επομένως απορρίπτεται η μηδενική υπόθεση.

Υπάρχουν λοιπόν σαφείς ενδείξεις ότι η διαφήμιση άλλαξε το μερίδιο αγοράς.

ΚΕΦΑΛΑΙΟ 2: Γεωγραφικά Συστήματα Πληροφοριών

2.1 Γενικά

Τα Γεωγραφικά Συστήματα Πληροφοριών (ΓΣΠ), γνωστά ευρέως και ως G.I.S. Geographic Information Systems, είναι συστήματα διαχείρισης χωρικών δεδομένων (spatial data) και συσχετισμένων ιδιοτήτων. Ένα ψηφιακό σύστημα είναι ικανό να ενσωματώσει, αποθηκεύσει, προσαρμόσει, αναλύσει και παρουσιάσει γεωγραφικά συσχετισμένες (geographically-referenced) πληροφορίες. Σε πιο γενική μορφή, ένα ΓΣΠ είναι ένα εργαλείο "έξυπνου χάρτη", το οποίο επιτρέπει στους χρήστες του να αποτυπώσουν μια περίληψη του πραγματικού κόσμου, να δημιουργήσουν ερωτήσεις χωρικού ή περιγραφικού χαρακτήρα (αναζητήσεις δημιουργούμενες από τον χρήστη), να αναλύσουν τα χωρικά δεδομένα (spatial data), να τα προσαρμόσουν και να τα αποδώσουν σε αναλογικά μέσα (εκτυπώσεις χαρτών και διαγραμμάτων) ή σε ψηφιακά μέσα (αρχεία χωρικών δεδομένων, διαδραστικοί χάρτες στο Διαδίκτυο).

Τα συστήματα GIS, αποτυπώνουν χωρικά δεδομένα σε γεωγραφικό ή χαρτογραφικό ή καρτεσιανό σύστημα συντεταγμένων. Βασικό χαρακτηριστικό τους είναι ότι τα χωρικά δεδομένα συνδέονται και με περιγραφικά δεδομένα, π.χ. μια ομάδα σημείων που αναπαριστούν θέσεις πόλεων συνδέεται με ένα πίνακα όπου κάθε εγγραφή εκτός από τη θέση περιέχει πληροφορίες όπως ονομασία, πληθυσμός κλπ.

Τα Συστήματα Γεωγραφικών Πληροφοριών είναι πληροφοριακά συστήματα (Information Systems) που παρέχουν την δυνατότητα συλλογής, διαχείρισης, αποθήκευσης, επεξεργασίας, ανάλυσης και οπτικοποίησης, σε ψηφιακό περιβάλλον, των δεδομένων που σχετίζονται με τον χώρο. Τα δεδομένα αυτά συνήθως λέγονται γεωγραφικά ή χαρτογραφικά ή χωρικά (spatial) και μπορεί να συσχετίζονται με μια σειρά από περιγραφικά δεδομένα τα οποία και τα χαρακτηρίζουν μοναδικά.

Η χαρακτηριστική δυνατότητα που παρέχουν τα GIS είναι αυτή της σύνδεσης της χωρικής με την περιγραφική πληροφορία (η οποία δεν έχει από μόνη της χωρική υπόσταση). Η τεχνολογία που χρησιμοποιείται για την λειτουργία αυτή βασίζεται:

- Είτε στο σχεσιακό (relational) μοντέλο δεδομένων, όπου τα περιγραφικά δεδομένα πινακοποιούνται χωριστά και αργότερα συσχετίζονται με τα χωρικά δεδομένα μέσω κάποιων μοναδικών τιμών που είναι κοινές και στα δύο είδη δεδομένων.
- Είτε στο αντικειμενοστραφές (object-oriented) μοντέλο δεδομένων, όπου τόσο τα χωρικά όσο και τα περιγραφικά δεδομένα συγχωνεύονται σε αντικείμενα, τα οποία μπορεί να μοντελοποιούν κάποια αντικείμενα με φυσική υπόσταση (π.χ. κατηγορία = "δρόμος", όνομα = "Πανεπιστημίου", γεωμετρία = "[X1,Y1],[X2,Y2]...", πλάτος = "20μέτρα").

Το αντικειμενοστραφές μοντέλο τείνει να χρησιμοποιείται όλο και περισσότερο σε εφαρμογές GIS εξαιτίας των αυξημένων δυνατοτήτων του σε σχέση με το σχεσιακό μοντέλο της δυνατότητας που παρέχει για την εύκολη και απλοποιημένη μοντελοποίηση σύνθετων φυσικών φαινομένων και αντικειμένων με χωρική διάσταση.

Πολλές φορές η ολοκληρωμένη έννοια των GIS (integrated GIS concept) επεκτείνεται για να συμπεριλάβει τόσο τα δεδομένα (που αποτελούν ουσιαστικά τον πυρήνα τους), το λογισμικό και τον μηχανικό εξοπλισμό, όσο και τις διαδικασίες και το ανθρώπινο δυναμικό, που αποτελούν αναπόσπαστα τμήματα ενός οργανισμού, ο οποίος έχει σαν πρωταρχική του δραστηριότητα την διαχείριση πληροφορίας με την βοήθεια GIS.

2.2 Τα μέρη ενός ΓΣΠ

Τα ΓΣΠ έχουν τρία βασικά συστατικά τα οποία είναι: ο απαραίτητος τεχνικός εξοπλισμός (Hardware), το λογισμικό (Software), και τα διαθέσιμα (Resourceware)

Τεχνικός Εξοπλισμός (Hardware)

Ο τεχνικός εξοπλισμός αποτελείται από τρία μέρη: τη κεντρική μονάδα (CPU), τα περιφερειακά και το τερματικό (DVU). Η κεντρική μονάδα επεξεργασίας ενός ΓΣΠ δε διαφέρει σε τίποτα από οποιουδήποτε άλλου γενικής χρήσης συστήματος πληροφοριών ενώ τα πιο γνωστά λειτουργικά συστήματα χρησιμοποιούνται από τα διάφορα πακέτα ΓΣΠ. Τα περιφερειακά διακρίνονται σε αυτά της εισόδου, που επιτρέπουν την είσοδο των στοιχείων (ψηφιοποιητές, σαρωτές) της εξόδου που συμμετέχουν στη παρουσίαση των στοιχείων (οθόνες, σχεδιαστές) και της διαχείρισης που βοηθούν στην αποθήκευση και διαχείριση των στοιχείων.

Λογισμικό (Software)

Το λογισμικό ενός ΓΣΠ περιλαμβάνει έξι βασικά "υποσυστήματα" για:

- Εισαγωγή δεδομένων και έλεγχο - επαλήθευση τους που καλύπτει τις ανάγκες μετασχηματισμού των στοιχείων από την αρχική τους μορφή (χάρτες, τηλεπισκοπικά προϊόντα κλπ) σε αναγνωρίσιμη ψηφιακή μορφή.
- Διαχείριση της γεωγραφικής βάσης δεδομένων, που αναφέρεται στο τρόπο που δομούνται και οργανώνονται τα χωρικά και μη χωρικά δεδομένα.
- Εξαγωγή πληροφοριών και παρουσίαση, που εστιάζεται στην παρουσίαση στοιχείων και αποτελεσμάτων των αναλυτικών διαδικασιών.
- Μετασχηματισμό των δεδομένων που στοχεύουν στο συντονισμό και την ανάλυση των στοιχείων.
- Αναζήτηση, που βοηθούν το χρήστη να επικοινωνεί με τον Η/Υ αναζητώντας λύση μέσα από μια σειρά ερωτήσεων.
- Εργαλεία χωρικής ανάλυσης, απαραίτητα για την κάλυψη αναγκών σε εφαρμογές που αναφέρονται στην ανάλυση χώρου.

Γενικά τα ΓΣΠ θα μπορούσαν να παραλληλισθούν με μια εργαλειοθήκη. Μερικά από τα εργαλεία ή συνδυασμοί τους μπορούν να χρησιμοποιηθούν για να επιλύσουν αρκετά χωρικά προβλήματα. Σύμφωνα με τον P.Burrough τα κυριότερα χωρικά προβλήματα που επιλύει ένα ΓΣΠ είναι τα εξής:

- Τι είναι στη θέση A;
- Πού είναι τα X1 και X2;
- Πού είναι το A σε σχέση με το B;
- Πόσο μεγάλο είναι το A (εμβαδόν, περίμετρος);
- Ποιο είναι το αποτέλεσμα της αλληλεπίθεσης μεταξύ δύο ή περισσότερων επιπέδων (overlay);
- Πόσες εμφανίσεις τύπου A βρίσκονται σε απόσταση ρ από το B;
- Προσδιορισμός της ζώνης σε απόσταση ρ από το A (buffering).

Τα κυριότερα εργαλεία που είναι ενσωματωμένα στα περισσότερα ΓΣΠ παραμένουν η αλληλεπίθεση και η δημιουργία ζώνης. Σε ορισμένα και καθώς εξελίσσονται συναντώνται περισσότερα και πιο ευέλικτα εργαλεία.

Διαθέσιμα (Resourceware)

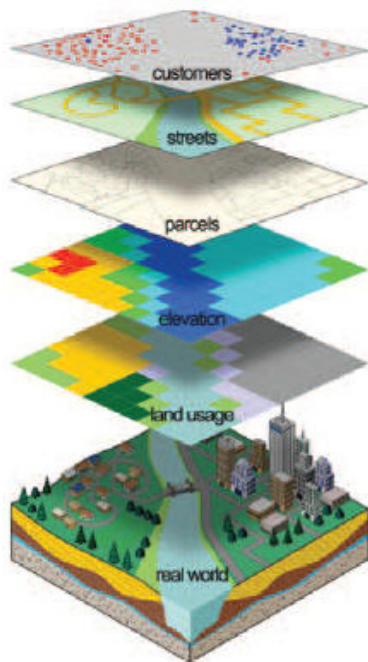
Τα διαθέσιμα παίζουν καθοριστικό ρόλο στην εγγύηση ότι οι διαδικασίες των ΓΣΠ είναι οι πλέον κατάλληλες και αποδοτικές. Ένα υπολογιστικό σύστημα με το αναγκαίο λογισμικό δεν εξασφαλίζει καμία επιτυχία σε οποιαδήποτε περίπτωση αν δεν υπάρχουν τα κατάλληλα στοιχεία – δεδομένα.

Τέλος, για να λειτουργήσει ένα σύστημα γεωγραφικών πληροφοριών χρειαζόμαστε περισσότερα από το υλικό μέρος, το λογισμικό και τα δεδομένα. Πρέπει να διαθέτουμε και το έμπυχο υλικό, δηλαδή τους ανθρώπους ή **χειριστές** που είναι απαραίτητοι για την λειτουργία του.

2.3 Δομές Δεδομένων

Σε ένα ΣΓΠ τα χωρικά δεδομένα μπορούν να αναπαρίστανται με δύο βασικές δομές: την διανυσματική δομή και τη ψηφιδωτή δομή. Σε όλα τα ΣΓΠ οι δύο δομές αποδίδονται ταυτόχρονα σε κοινές απεικονίσεις ενώ πολλά λογισμικά GIS προσφέρουν την δυνατότητα μετάβασης από τη μία δομή στην άλλη.

Ένας συνήθης τρόπος δόμησης γεωγραφικών και χαρτογραφικών δεδομένων είναι σε **θεματικά επίπεδα** (layers). Κάθε επίπεδο αντιπροσωπεύοντας μια θεματική προσέγγιση για ένα συγκεκριμένο σκοπό ή ομάδα αναγκών, μπορεί να περιλαμβάνει ένα ή περισσότερα είδη πληροφοριών.



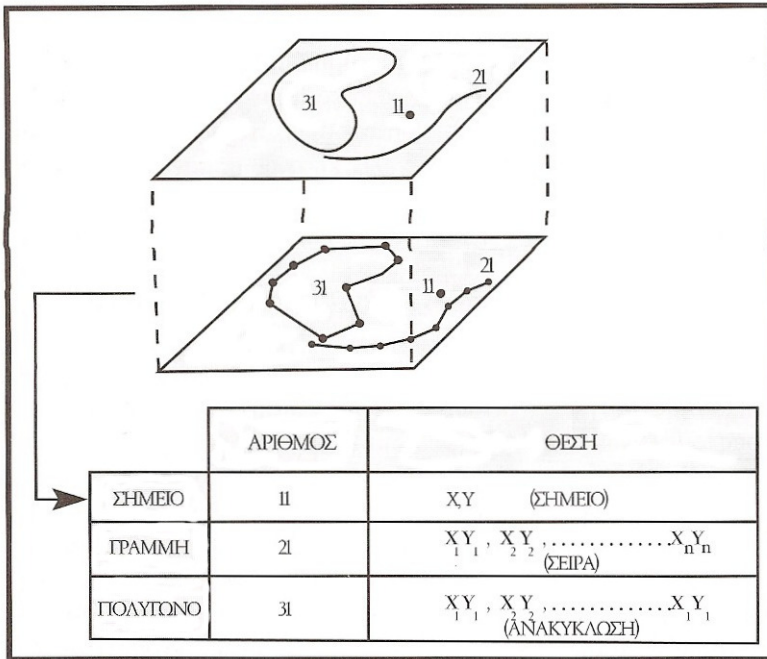
σχήμα 2.1: οργάνωση σε layers (θεματικά επίπεδα)

Όσον αφορά τις δύο δομές δεδομένων έχουμε:

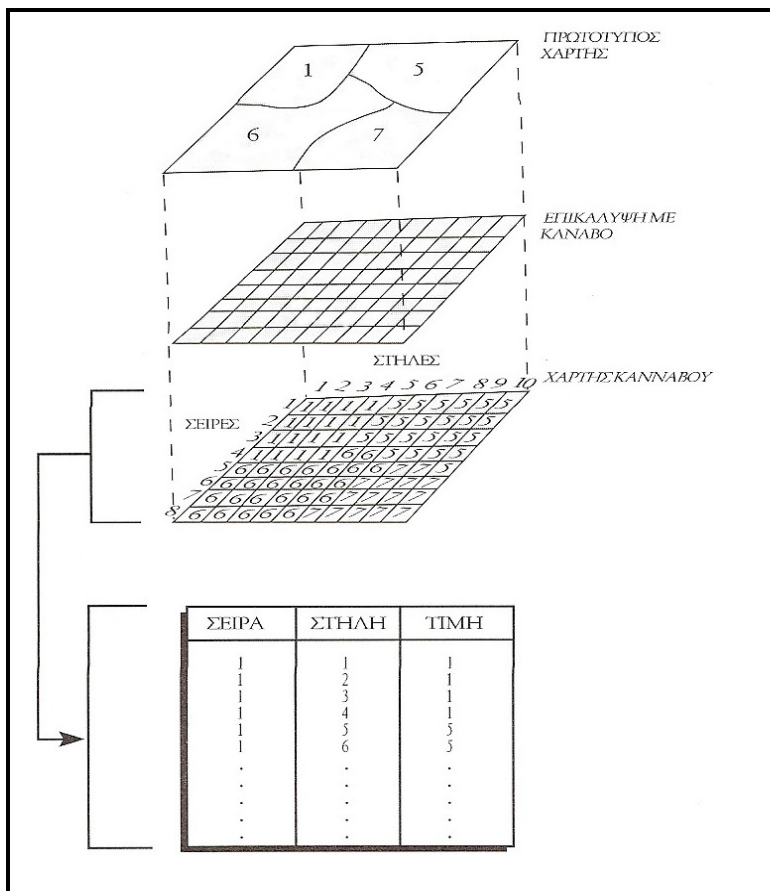
Διανυσματική δομή (vector data structure): Τα χωρικά δεδομένα αναπαρίστανται με τρεις βασικούς τύπους γεωμετριών, **σημεία, γραμμές, πολύγωνα**. Η χρήση του κάθε χωρικού αρχετύπου εξαρτάται από τη κλίμακα παρατήρησης ή το επίπεδο γενίκευσης. Σε μικρή κλίμακα, οι πόλεις αναπαρίστανται από σημεία, ενώ οι δρόμοι και τα ποτάμια από γραμμές. Όμως καθώς αυξάνεται η κλίμακα παρατήρησης, πρέπει να λαμβάνεται υπόψη και η επιφανειακή διάσταση των φαινομένων. Έτσι σε μεγαλύτερες κλίμακες, η πόλη αναπαρίσταται από ένα πολύπλοκο σύνολο χωρικών αρχετύπων που ορίζουν τα όρια των κτιρίων, τους δρόμους, τα πάρκα και άλλα φυσικά και τεχνητά φαινόμενα από τα οποία αποτελείται.

Οι γραμμές ή τα τόξα είναι σειρές διατεταγμένων σημείων των οποίων καταγράφονται οι συντεταγμένες. Οι επιφάνειες ή τα πολύγωνα αποθηκεύονται επίσης ως διατεταγμένες σειρές σημείων, όμως το αρχικό και το τελικό σημείο ταυτίζονται έτσι ώστε να ορίζεται ένα κλειστό σχήμα. Τα σημεία σε μια διανυσματική δομή δεδομένων περιγράφονται ως γραμμές μηδενικού μήκους.

Κανονικοποιημένη δομή δεδομένων (raster data structure): Αντιλαμβάνεται το χώρο ως μωσαϊκή διαίρεση φατνίων – ψηφίδων (tessellation of cells), δηλαδή τοποθετεί και αποθηκεύει τα δεδομένα χρησιμοποιώντας ένα πίνακα ή κάναβο φατνίων. Το φατνίο συνήθως αναφέρεται ως εικονοστοιχείο (pixel). Η κατηγορία των κανονικοποιημένων χωρικών δομών δεν περιλαμβάνει μόνο τις δομές που βασίζονται σε ένα ορθογώνιο ή τετραγωνικό πλέγμα. Στη πραγματικότητα αυτή η κατηγορία περιλαμβάνει κάθε συνεχώς επαναλαμβανόμενο πρότυπο ενός κανονικού πολυγώνου ή πολυέδρου. Πιο συγκεκριμένα οι τρεις τύποι κανονικών διαιρέσεων που χρησιμοποιούνται στα χωρικά μοντέλα δεδομένων οι **τετραγωνικές, οι τριγωνικές και οι εξαγωνικές**. Καθεμία από αυτές έχει διαφορετικά λειτουργικά χαρακτηριστικά που βασίζονται στη διαφορετική γεωμετρία των στοιχειωδών πολυγώνων.



σχήμα 2.2: διανυσματική δομή δεδομένων



σχήμα 2.3: κανονικοποιημένη δομή δεδομένων

Τοπολογία

Η τοπολογία αποτελεί ένα σύνολο γεωμετρικών κανόνων που πρέπει να ακολουθεί η γεωμετρική πληροφορία ανάλογα με τη φύση της, για παράδειγμα όμορα αντικείμενα, όπως δύο γειτονικοί νομοί της χώρας μοιράζονται κοινό όριο. Η τοπολογική δομή στο διανυσματικό σύστημα προσδιορίζει με ακρίβεια πως και που τα σημεία και οι γραμμές ενός χάρτη συνδέονται μέσω των **κόμβων** (τοπολογικά σημεία συνένωσης). Η σειρά της συνδεσιμότητας ορίζει το σχήμα ενός τόξου ή ενός πολυγώνου.

2.4 Πηγές Δεδομένων για τα ΓΣΠ

Τα γεωγραφικά δεδομένα μπορούν να ταξινομηθούν σε δύο κατηγορίες:

- Δεδομένα για το περιβάλλον και τα φυσικά διαθέσιμα [Θεματικά, Τοπογραφικά]
- Κοινωνικο-οικονομικά δεδομένα

Η συλλογή των δεδομένων μπορεί να προέρχεται από πρωτογενείς ή δευτερογενείς πηγές εκ των οποίων οι πιο σημαντικές συνοπτικά είναι:

- Τοπογραφικές μετρήσεις
- Κτηματολόγιο
- Αεροφωτογραφίες
- Δορυφορικές εικόνες
- Υπάρχοντες χάρτες - Ψηφιοποίηση
- Διαθέσιμα ψηφιακά δεδομένα

2.5 Επεξεργασία δεδομένων – Χωρική ανάλυση

Η χωρική ανάλυση μπορεί να θεωρηθεί ως μια γενική ικανότητα να χειριστεί κανείς τη χωρική πληροφορία σε διαφορετικές μορφές με αποτέλεσμα να εξάγει επιπρόσθετη σημασία. Ουσιαστικά κατά τη διαδικασία της χωρικής ανάλυσης ο χρήστης διατυπώνει κάποιο πρόβλημα ή ερωτήματα. Η βάση δεδομένων περιέχει πληροφορία που μπορεί να χρησιμοποιηθεί για την επίλυση του προβλήματος. Η διαδικασία αυτή μετατρέπει τα αρχικά δεδομένα της βάσης σε παράγωγα δεδομένα ή παράγωγο χάρτη, πίνακα ή σχήμα που θα παρέχει την απάντηση στην ερώτηση του χρήστη.

Στόχοι της χωρικής ανάλυσης συνοψίζονται παρακάτω:

- Ακριβής περιγραφή των γεγονότων στο γεωγραφικό χώρο.
- Συστηματική διερεύνηση των χωρικών προτύπων των γεγονότων και συσχέτιση των γεγονότων στο χώρο.
- Βελτίωση της ικανότητας πρόβλεψης και ελέγχου των γεγονότων που προκύπτουν στο γεωγραφικό χώρο.

Τα ΓΣΠ παρέχουν δυνατότητες για μετασχηματισμό των αρχικών δεδομένων και απάντηση ερωτήσεων που διατυπώνουν οι χρήστες. Μερικοί τέτοιοι μετασχηματισμοί είναι κοινοί με αντίστοιχους που γίνονται σε σχεδιαστικά προγράμματα CAD (διαγραφή, ενημέρωση δεδομένων, αλλαγή κλίμακας προβολής). Τα ΓΣΠ παρέχουν επιπλέον ένα ευρύ φάσμα αναλυτικών διαδικασιών που αφορούν τη τοπολογία, τα χωρικά χαρακτηριστικά των γεωγραφικών δεδομένων, τα μη χωρικά χαρακτηριστικά.

Παρακάτω παρουσιάζονται επιγραμματικά – μιας και η αναλυτική παρουσίαση τους ξεφεύγει από το πλαίσιο της γενικής παρουσίασης των ΓΣΠ - οι βασικές τεχνικές διαχείρισης και ανάλυσης των χωρικών δεδομένων.

- **Ανάκτηση δεδομένων** μέσω διατύπωσης ερωτημάτων και χρήση λογικών τιμών 0 ή 1 στην αποθηκευμένη πληροφορία.
- **Αφαιρετική διαδικασία**, είναι παρόμοια με αυτή της γενίκευσης αλλά υλοποιείται διαφορετικά (υπολογισμός κεντροειδών, ισαριθμικές γραμμές κ.α.).
- **Διαχείριση φύλλων χάρτη** (πχ μεταβολή κλίμακας, αλλαγή χαρτογραφικής προβολής, μετασχηματισμοί συντεταγμένων κ.α.).
- **Δημιουργία ζωνών επιρροής**, δηλαδή τη δημιουργία νέων πολυγώνων γύρω από σημειακές, γραμμικές, επιφανειακές οντότητες, με κάποιο εύρος.
- **Επίθεση και διαχωρισμός πολυγώνων**, περιλαμβάνει τη σύνθεση ή τη διάσπαση πολλαπλών χαρτών (δύο ή περισσότερων) ώστε να παραχθεί μια νέα ομάδα δεδομένων.
- **Μετρήσεις**, που αφορούν σημεία, γραμμές, επιφάνειες και όγκους.
- **Ανάλυση δικτύου**, με την έννοια δίκτυο νοείται ένα σύνολο αλληλοσυνδεδεμένων γραμμών που σχηματίζουν ένα σύνολο στοιχείων μέσα από τα οποία ρέουν πόροι. (προβλήματα βέλτιστης διαδρομής κ.α.).
- **Ανάλυση πινακοποιημένων δεδομένων**
- **Ανάλυση ψηφιακού ανάγλυφου**, που περιλαμβάνει τον υπολογισμό διαφόρων προϊόντων από ένα ψηφιακό μοντέλο υψομέτρων (D.E.M.)

Κύριες μορφές προϊόντων από ένα ΓΣΠ (έξοδος) είναι οι αναλογικοί χάρτες, οι διάφοροι πίνακες με στατιστικά δεδομένα, η εμφάνιση στην οθόνη γραφικών και μη γραφικών χαρακτηριστικών και τα διάφορα αρχεία που προκύπτουν από την επεξεργασία των γεωγραφικών δεδομένων

ΚΕΦΑΛΑΙΟ 3 : ΕΦΑΡΜΟΓΗ

3.1 Γενικά

Στο κεφάλαιο αυτό παρουσιάζεται αναλυτικά το εκπαιδευτικό υλικό που αναπτύχθηκε για το μάθημα ' Γεωγραφία και Ανάλυση Χώρου ', που διδάσκεται στη ΣΑΤΜ.

Συγκεκριμένα αναπτύχθηκε μία άσκηση εφαρμογής πέντε βασικών δειγματοληπτικών μεθόδων σε περιβάλλον γεωγραφικού συστήματος πληροφοριών, σε χωρικό υπόβαθρο, που αποτελεί έναν διανυσματικό χάρτη χρήσεων γης μιας περιοχής. Συγκεκριμένα εφαρμόστηκαν οι μέθοδοι:

- της απλής τυχαίας σημειακής χωρικής δειγματοληψίας,
- της συστηματικής σημειακής χωρικής δειγματοληψίας,
- της στρωματοποιημένης σημειακής χωρικής δειγματοληψίας,
- της τυχαίας γραμμικής δειγματοληψίας,
- και της συστηματικής γραμμικής δειγματοληψίας

Αντικείμενο της δειγματοληψίας είναι η εκτίμηση των ποσοστών των επιμέρους χρήσεων γης που παρουσιάζονται σε χάρτη.

Οι τρεις σημειακές μέθοδοι πραγματοποιούνται σε ψηφιακό χάρτη με τη βοήθεια μιας γεννήτριας τυχαίων σημείων. Ουσιαστικά η γεννήτρια τυχαίων σημείων παράγει τυχαίες συντεταγμένες στα όρια της έκτασης που υποδεικνύει ο χρήστης και δημιουργεί νέο θεματικό επίπεδο (σημειακό), με στοιχεία τον αριθμό των τυχαίων σημείων που επιλέχθηκε να δημιουργηθούν.

Οι γραμμικές μέθοδοι πραγματοποιούνται με λίγο πιο σύνθετες διαδικασίες που περιγράφονται αναλυτικά στις αντίστοιχες παραγράφους.

Η αυξανόμενη ανάγκη για διαχείριση, ανάλυση, επεξεργασία και παρουσίαση των διαφόρων γεωγραφικών δεδομένων, έχει κάνει τη χρήση των GIS (geographic information systems) απαραίτητη στις περισσότερες γεωγραφικές μελέτες.

Τα γεωγραφικά συστήματα πληροφοριών αποτελούν ένα ιδανικό μέσο και για τη διδασκαλία θεμάτων γεωγραφίας δίνοντας τη δυνατότητα στους σπουδαστές να αναπτύξουν σημαντικές δεξιότητες στην επίλυση προβλημάτων. Τα προηγούμενα χρόνια περιορίζονταν σε επαγγελματίες με πρόσβαση σε υψηλού επιπέδου λογισμικό και υπολογιστικό εξοπλισμό. Τα Γ.Σ.Π. είναι πλέον προσιτά και χρησιμοποιούνται ευρέως.

Η εφαρμογή σχεδιάστηκε στο πρόγραμμα ArcGIS 9.2.

Η χρήση ΓΣΠ στην εφαρμογή, εκτός του ότι συμβάλει στη καλύτερη και πιο φιλική παρουσίαση με στόχο την κατανόηση των βασικών εννοιών της δειγματοληψίας μπορεί να θεωρηθεί ότι βοηθά:

- Στην εξοικείωση των σπουδαστών με ένα σύγχρονο λογισμικό πακέτο ΓΣΠ,
- Στην ανάπτυξη ικανότητας ερμηνείας χαρτών και στην επέμβαση σε αυτούς με τα διάφορα εργαλεία που προσφέρει το λογισμικό,
- Στη χρήση διαφόρων χαρτογραφικών συμβολισμών και στον πειραματισμό με αυτούς, ώστε να προκύψει ένα κατά το δυνατό βέλτιστο οπτικά προϊόν,

- Στην εισαγωγή των σπουδαστών στην ανάλυση του χώρου μέσα από ρεαλιστικά παραδείγματα και στην αντιμετώπιση των διαφόρων προβλημάτων που προκύπτουν.

Τέλος με τη χρήση ΓΣΠ για τη διενέργεια των δειγματοληψιών μειώνεται στο ελάχιστο ο χρόνος εφαρμογής (πχ. παραγωγή τυχαίων σημείων), επιτρέποντας στους σπουδαστές να εμβαθύνουν περισσότερο στο θεωρητικό κομμάτι αυτών των διαδικασιών.

Ένας αναλογικός σχεδιασμός της εφαρμογής, δηλαδή η πραγματοποίηση δειγματοληψιών πάνω σε αναλογικό χάρτη θα έθετε σημαντικούς περιορισμούς.

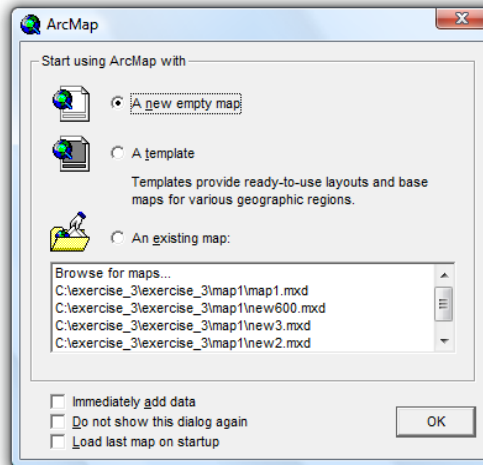
Αρχικά το μέγεθος του δείγματος δε θα μπορούσε να ξεπερνά τις 20 με 30 παρατηρήσεις (σημεία), έτσι δεν θα ήταν δυνατό να υπάρχει – σε εύλογο χρονικό διάστημα – η ακρίβεια των δειγματοληψιών που επιτυγχάνεται με τις παρακάτω διαδικασίες.

Τα τυχαία σημεία πάνω σε έναν αναλογικό χάρτη θα τοποθετούνταν ένα – ένα με τη βοήθεια πίνακα τυχαίων αριθμών και προφανώς ο υπολογισμός των σημείων που αντιστοιχούν σε κάθε κατηγορία του πειράματος μας θα γίνονταν πάλι χειροκίνητα, με πιθανότητα πάντα χονδροειδών σφαλμάτων (σφάλματα καταγραφής κ.α.).

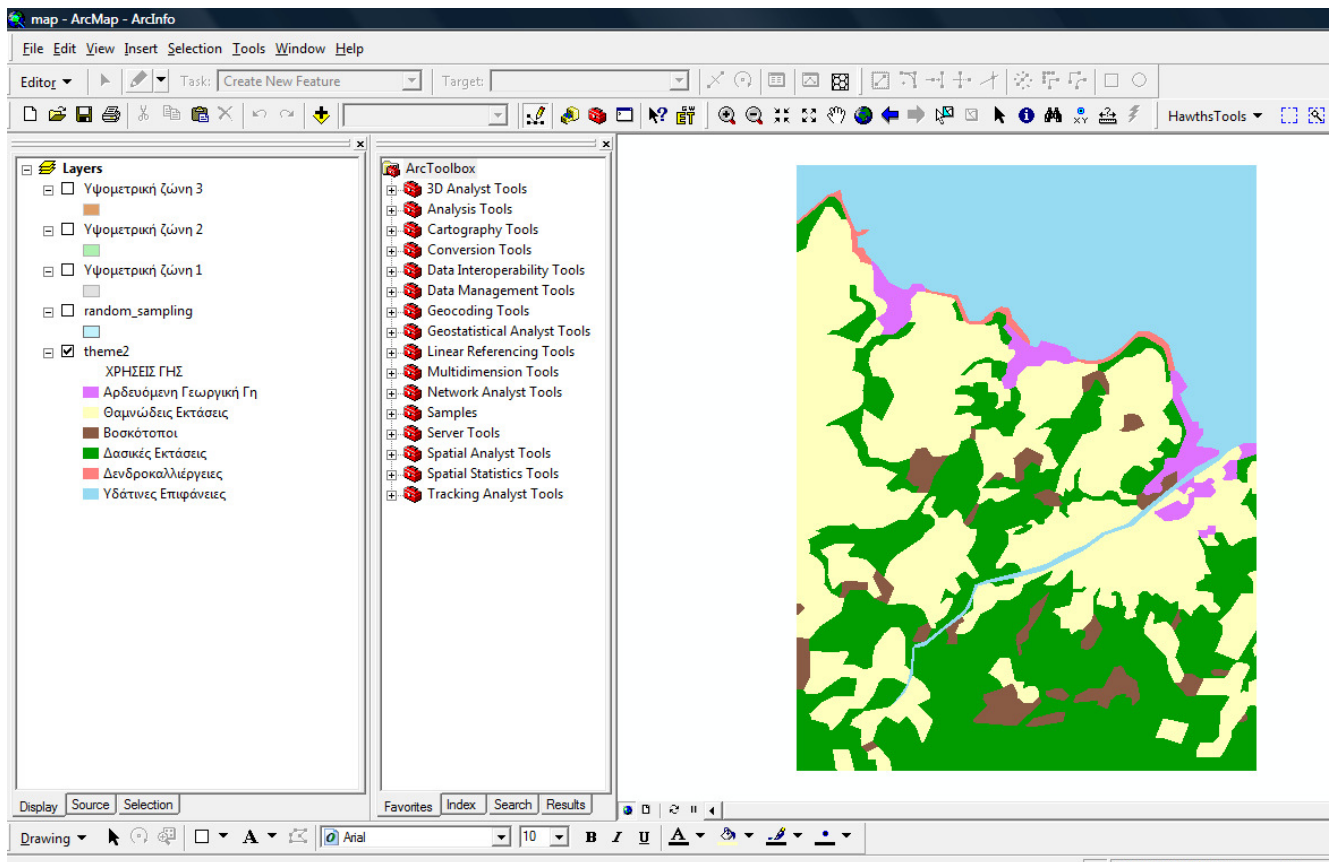
Παράλληλα θα έπρεπε να υπάρχουν υπολογισμένα ήδη τα πραγματικά - αναμενόμενα ποσοστά της κάθε κατηγορίας, ώστε να γίνει έλεγχος καλής προσαρμογής.

3.2 Αναλυτικά βήματα Εφαρμογής μεθόδων δειγματοληψίας

Αρχικά, γίνεται εκκίνηση του arcmap [έναρξη > προγράμματα > arcgis > arcmap]. Στο παράθυρο που εμφανίζεται επιλέγεται εκκίνηση του arcmap με ένα νέο άδειο χάρτη κάνοντας κλικ στο OK.



Για να ανοίξει ο χάρτης της εφαρμογής [file > open > (πηγαίνοντας στη θέση του φακέλου που βρίσκονται τα δεδομένα της άσκησης) > map (σε μορφή arcmap document)]. Έτσι στην οθόνη εμφανίζεται πλέον ο χάρτης. (σχήμα 3.1)



σχήμα 3.1: Περιβάλλον Arcmap – Αρχικά δεδομένα

Ο χάρτης αποτελείται από μια σειρά θεματικών επιπέδων (layers) που βρίσκονται στον πίνακα περιεχομένων, το επίπεδο [**theme2**] που περιλαμβάνει τα πολύγωνα των 5 χρήσεων γης που θα πάρουν μέρος στη δειγματοληψία και τις υδάτινες περιοχές (1 πολύγωνο) που δεν συμμετέχουν. Το επίπεδο [**random_sampling**] που είναι το υπόβαθρο για την δοκιμαστική και την απλή τυχαία δειγματοληψία (1 πολύγωνο). Τα επίπεδα [**Υψομετρική ζώνη 1**], [**Υψομετρική ζώνη 2**], [**Υψομετρική ζώνη 3**] περιλαμβάνουν τα αντίστοιχα πλαίσια που θα χρησιμοποιηθούν για την διενέργεια της στρωματοποιημένης δειγματοληψίας και το επίπεδο [**systematic_sampling**], που είναι ένα πολυγωνικό σχηματικό αρχείο και θα δημιουργήσουμε το πλαίσιο της συστηματικής τυχαίας δειγματοληψίας.

Στον ψηφιακό χάρτη [theme_2] απεικονίζονται οι παρακάτω χρήσεις γης:

Αρδευόμενη Γεωργική Γη (Irrigating Agricultural Soil)
Θαμνώδεις Εκτάσεις (Bushy Area)
Βοσκότοποι (Grazing Ground)
Δασικές Εκτάσεις (Forest)
Δενδροκαλλιέργειες (Tree-Cultivation)
και οι υδάτινες περιοχές.

Το χαρτογραφικό υπόβαθρο ήταν σε μορφή Arcinfo interchange file και μετατράπηκε σε αρχείο δεδομένων για να εισαχθεί στο Arcmap. Από το αρχικό αρχείο διατηρήθηκαν μόνο τα στοιχεία και οι πληροφορίες των πολυγώνων, τα οποία και αποθηκεύτηκαν σε μορφή πολυγωνικού σχηματικού αρχείου (shapefile). Ένα σχηματικό αρχείο μπορεί να περιέχει μόνο ενός είδους γεωγραφικό συμβολισμό (σημεία, γραμμές ή πολύγωνα).

Το σύστημα αναφοράς είναι αυθαίρετο και οι συντεταγμένες παρουσιάζονται σε μέτρα. Στον πίνακα περιγραφικών χαρακτηριστικών του χάρτη [attributes of theme_2], υπάρχουν διάφορα πεδία, μεταξύ των οποίων η έκταση και η περίμετρος του κάθε πολυγώνου.

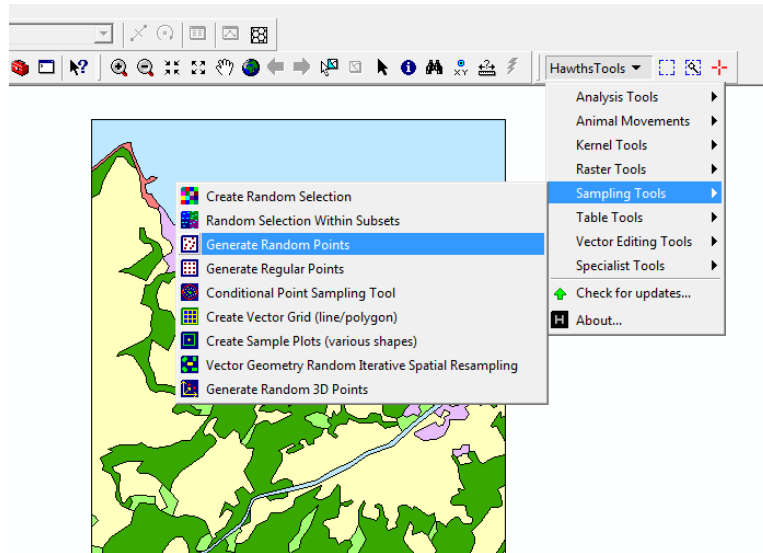
Παρουσίαση εργαλείων δειγματοληψίας της επέκτασης Hawth's Tools

Μετά την επιτυχή εγκατάσταση της επέκτασης, η ενεργοποίηση της εργαλειοθήκης γίνεται από το μενού επιλογών του arcmap και συγκεκριμένα από το μενού [**tools**].

Επιλέγοντας [**customize**], στην καρτέλα [**toolbars**] μαρκάρεται η επιλογή [**Hawth's Tools**] και έπειτα [**close**]. Εάν δεν έχει εμφανιστεί ήδη η νέα εργαλειοθήκη, κάνοντας δεξί κλικ στο γκρι πλαίσιο – όπου τοποθετούνται οι μπάρες εργαλείων, μαρκάρεται η επιλογή Hawth's Tools.

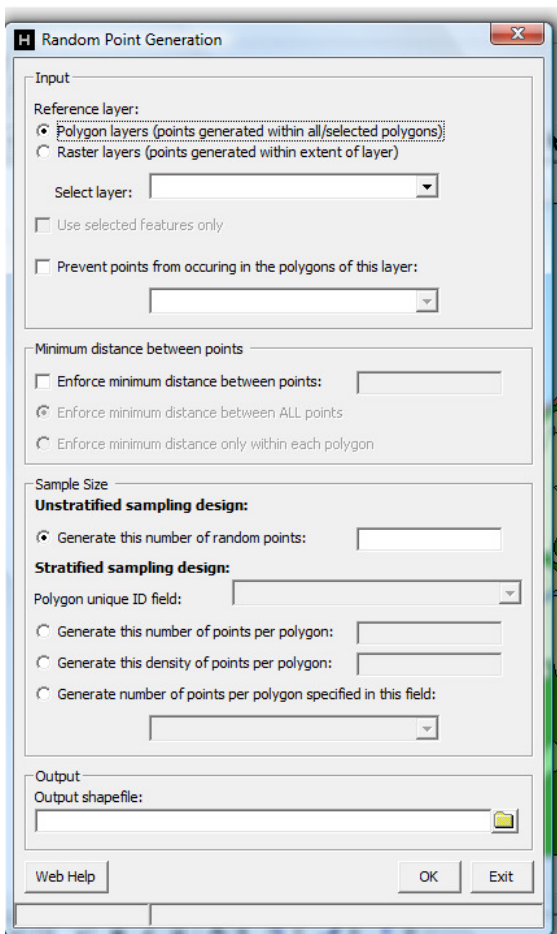
Πλέον, δίπλα σε κάποια από τις ήδη υπάρχουσες εργαλειοθήκες έχει τοποθετηθεί και η νέα μπάρα, έτοιμη για χρήση.

Η επέκταση (extension) Hawth's Tools, παρέχει μια σειρά από χρήσιμα εργαλεία μεταξύ των οποίων και εργαλεία δειγματοληψίας.



σχήμα 3.2: επιλογή γεννήτριας τυχαίων σημείων

Το πρώτο εργαλείο που θα χρησιμοποιηθεί στο πλαίσιο των μεθόδων δειγματοληψίας είναι η γεννήτρια τυχαίων σημείων [**sampling tools > generate random points**].



σχήμα 3.3: Παράθυρο διαλόγου Random Points Generator

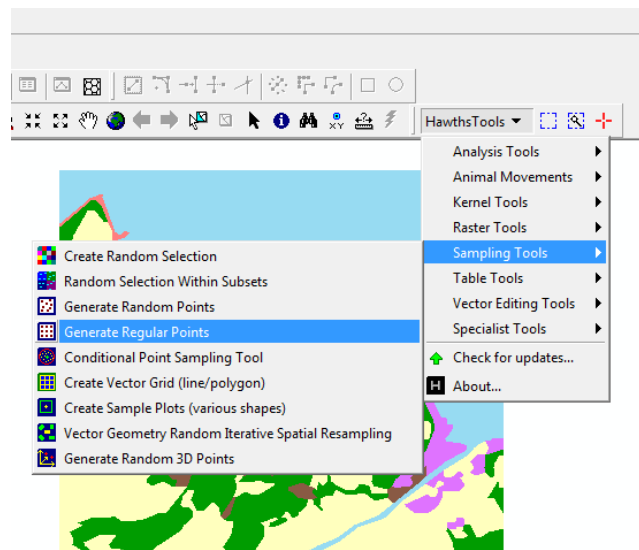
Στο διπλανό παράθυρο που εμφανίζεται επιλέγεται το επίπεδο αναφοράς [**reference layer**]

αν πρόκειται για πολυγωνικό ή raster και έπειτα επιλέγεται το επίπεδο πάνω στο οποίο θα τοποθετηθούν τα τυχαία σημεία. Ακολουθούν κάποιες επιλογές που δημιουργούν κανόνες για την τοποθέτηση των σημείων. (απαγόρευση σημείων να βρίσκονται σε πολύγωνα συγκεκριμένου επιπέδου και επιβολή ελάχιστης απόστασης μεταξύ των σημείων).

Έπειτα επιλέγεται ο αριθμός των σημείων που θα δημιουργηθούν δηλαδή το μέγεθος δείγματος που θα χρησιμοποιηθεί [**generate this number of random points**]. Ακολουθούν κάποιες επιλογές που επίσης δημιουργούν κανόνες σχετικά με τη τοποθέτηση των σημείων και επίσης δεν είναι απαραίτητο να συμπληρωθούν.

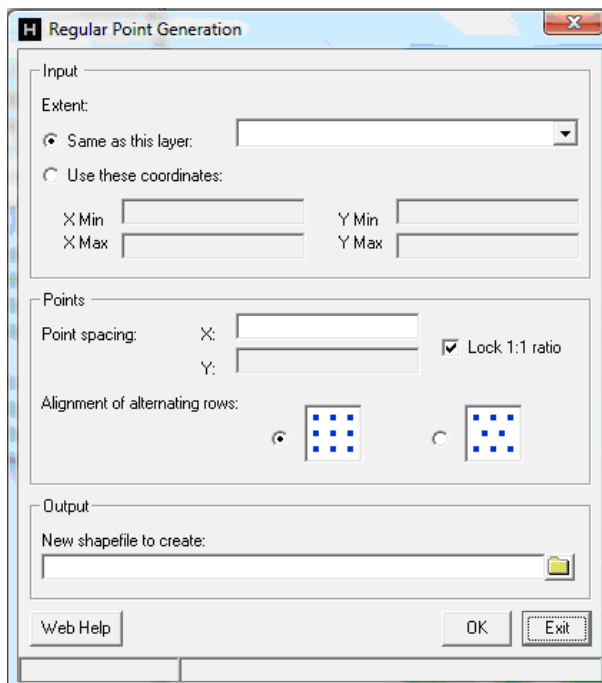
Τέλος, επιλέγεται το όνομα και η τοποθεσία του σχηματικού αρχείου που θα δημιουργηθεί.

Εκτός από το εργαλείο δημιουργίας τυχαίων σημείων, η μπάρα Hawth's tools, έχει ένα εργαλείο για τη δημιουργία κανονικά – ομοιόμορφα καταναμημένων σημείων. [generate regular points], που βρίσκεται και αυτό στο μενού [sampling tools].



σχήμα 3.4: επιλογή γεννήτριας ομοιόμορφα καταναμημένων σημείων

Επιλέγοντας τη δεύτερη αυτή γεννήτρια σημείων, εμφανίζεται ένα παράθυρο διαλόγου (σχήμα 3.5)



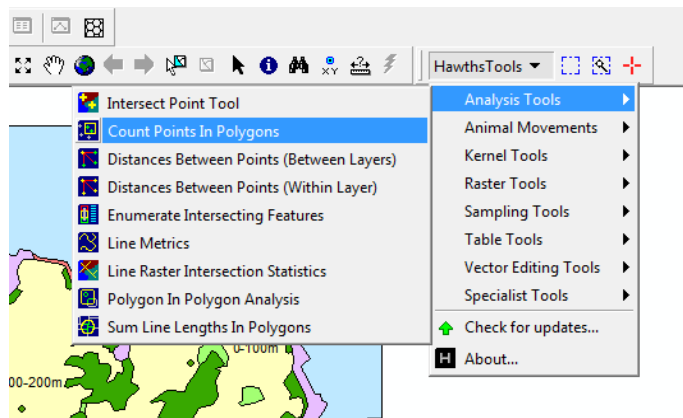
σχήμα 3.5: Παράθυρο διαλόγου regular point generation

Στο διπλανό παράθυρο που εμφανίζεται επιλέγεται η έκταση που θα τοποθετηθούν τα σημεία, είτε επιλέγοντας να είναι η ίδια με τα όρια κάποιου θεματικού επιπέδου, είτε θέτοντας οι ίδιοι τις συντεταγμένες.

Επόμενη επιλογή είναι το βήμα κατά X και Y ή point spacing, καθώς και ο τρόπος διάταξης των σημείων (δύο μορφές διάταξης).

Τέλος, επιλέγεται η θέση αποθήκευσης και το όνομα του σημειακού σχηματικού αρχείου που θα δημιουργηθεί.

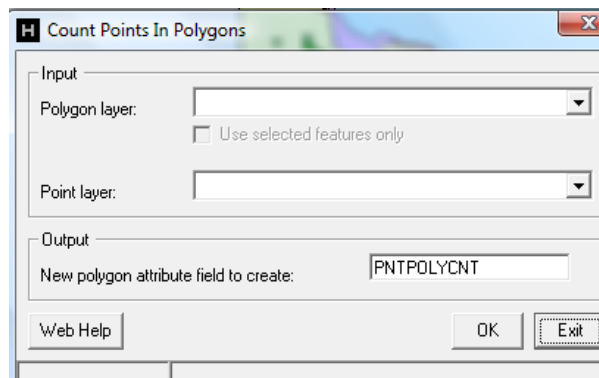
Ένα επίσης χρήσιμο εργαλείο που παρέχει η εργαλειοθήκη Hawth's Tools είναι ο υπολογισμός σημείων που βρίσκονται σε κάθε πολύγωνο του χάρτη - Count Points In Polygons, [Hawth Tools > Analysis Tools > Count Points In Polygons].



σχήμα 3.6: επιλογή count points in polygons

Στο παράθυρο διαλόγου που εμφανίζεται, επιλέγεται ως είσοδος το θεματικό επίπεδο που απαρτίζεται από τα πολύγωνα και το σημειακό επίπεδο.

Ως έξοδος επιλέγεται το όνομα της στήλης που θα δημιουργηθεί στον πίνακα περιγραφικών χαρακτηριστικών του πολυγωνικού θεματικού επιπέδου και για κάθε εγγραφή (πολύγωνο) θα αναγράφει τον αριθμό των σημείων που βρίσκονται επί αυτού.



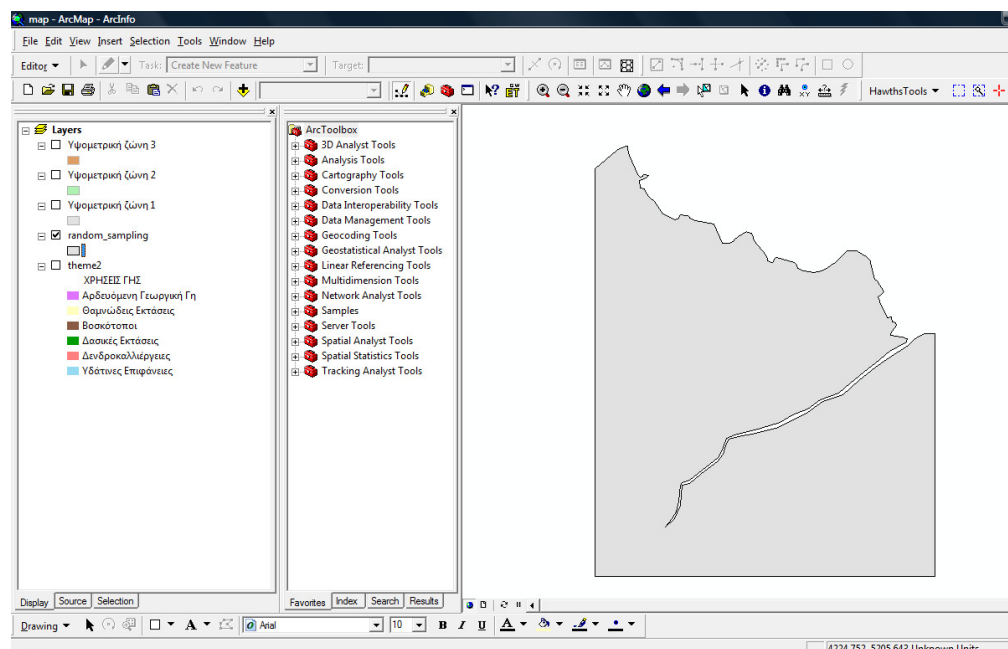
σχήμα 3.7: παράθυρο διαλόγου count points in polygons

3.3 Εφαρμογή Δοκιμαστικής Δειγματοληψίας

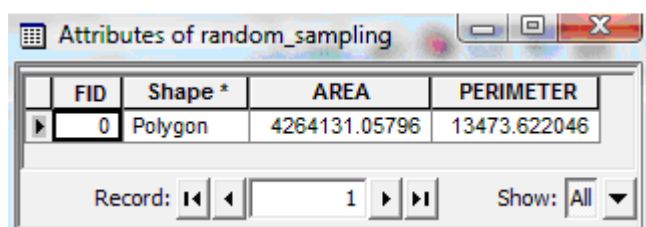
Αρχικό στάδιο της εφαρμογής των δειγματοληψιών είναι ο καθορισμός του μεγέθους του δείγματος. Για τον καθορισμό αυτό συνήθως συλλέγεται ένα δοκιμαστικό δείγμα που καλό είναι να περιλαμβάνει το λιγότερο 30 παρατηρήσεις.

Το δειγματοληπτικό υπόβαθρο για την απλή τυχαία δειγματοληψία είναι όλος ο χάρτης πλην των υδάτινων επιφανειών. Στον πίνακα περιεχομένων εμφανίζεται ως **'random_sampling'**. (σχήμα 3.8).

Το υπόβαθρο δημιουργήθηκε με διαδικασία επιλογής από το επίπεδο [theme_2] όλων των πολυγώνων πλην αυτού που απεικονίζει τις υδάτινες περιοχές. [select by attributes]. Επίσης έγινε απλοποίηση των γραμμών, ώστε να παρουσιάζεται ως ένα ενιαίο πολύγωνα. Στον πίνακα περιγραφικών χαρακτηριστικών του random_sampling, υπάρχει το εμβαδό και η περίμετρος του (σχήμα 3.9).



σχήμα 3.8: Δειγματοληπτικό υπόβαθρο απλής τυχαίας δειγματοληψίας

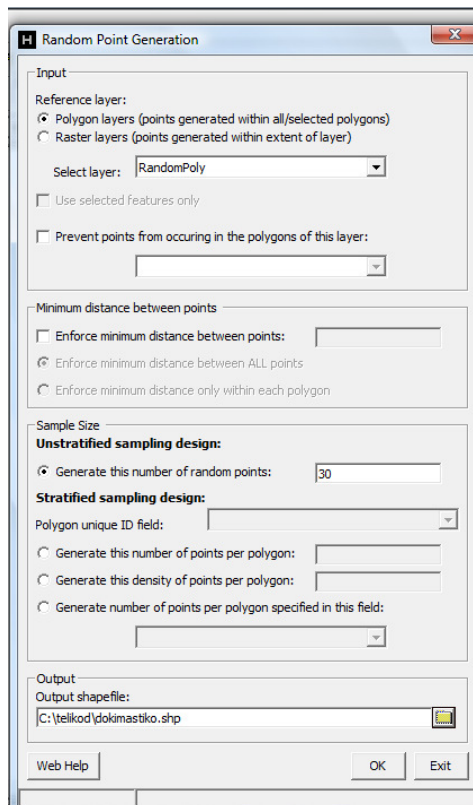


FID	Shape *	AREA	PERIMETER
0	Polygon	4264131.05796	13473.622046

σχήμα 3.9: πίνακας περιγραφικών χαρακτηριστικών του επιπέδου random_sampling (έκταση – περίμετρος).

Η εφαρμογή ξεκινά με τη λήψη ενός δοκιμαστικού δείγματος 30 παρατηρήσεων. Αυτό γίνεται με την επιλογή "generate random points" από τη εργαλειομπάρα Hawth's tools. [Hawths Tools > sampling tools > generate random points].

Το παράθυρο διαλόγου που εμφανίζεται συμπληρώνεται όπως παρακάτω:



Το θεματικό επίπεδο που επιλέγεται είναι το δειγματοληπτικό υπόβαθρο [RandomPoly].

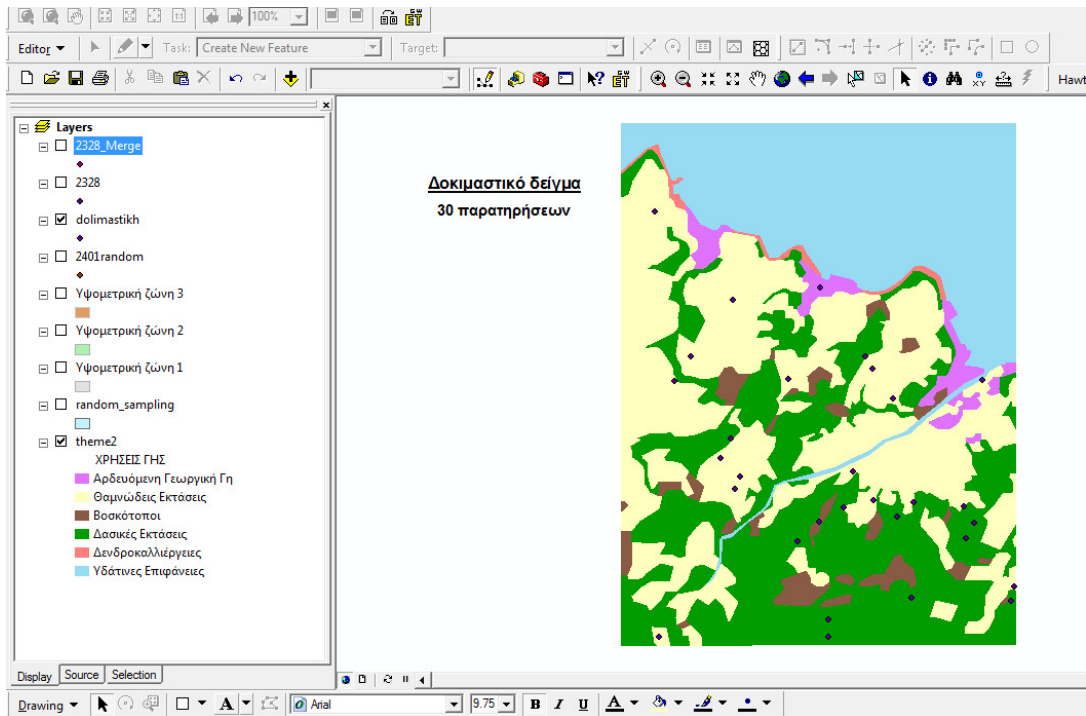
Ο αριθμός των σημείων που θα δημιουργηθούν επιλέγεται να είναι 30.

Τέλος δίνεται η θέση και το όνομα του σχηματικού αρχείου που θα δημιουργηθεί π.χ. [C:\deigma\dokimastiko.shp]

σχήμα 3.10: παράμετροι τυχαίας δοκιμαστικής δειγματοληψίας

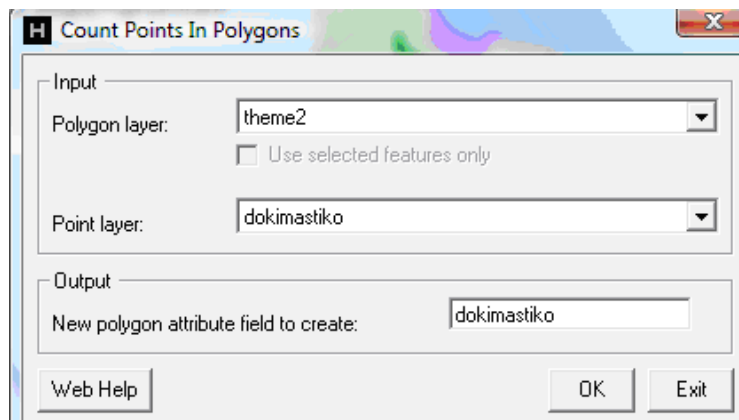
Το επίπεδο αυτό θα αποτελείται από 30 σημεία τυχαία κατανομημένα στο υπόβαθρο και κατ επέκταση στο επίπεδο χρήσεων γης .

Το σημειακό αυτό επίπεδο θα εμφανίζεται με το όνομα π.χ. "dokimastiko" στο πίνακα περιεχομένων.



σχήμα 3.11 Δοκιμαστικό δείγμα 30 παρατηρήσεων

Έπειτα υπολογίζουμε τον αριθμό των σημείων της δοκιμαστικής δειγματοληψίας που βρίσκονται σε κάθε πολύγωνο του θεματικού επιπέδου των χρήσεων γης [theme1], χρησιμοποιώντας το εργαλείο "count points in polygons". [Hawths Tools > Analysis Tools > Count points in polygons].



σχήμα 3.12: Υπολογισμός σημείων δοκιμαστικής ανά πολύγωνο του επιπέδου theme2 (επίπεδο χρήσεων γης).

Αυτό έχει ως αποτέλεσμα τη δημιουργία μιας νέας στήλης (dokimastiko) στον πίνακα περιγραφικών χαρακτηριστικών (attributes of theme2) του επιπέδου χρήσεων γης [theme2]. Για να ανοίξουμε τον πίνακα περιγραφικών χαρακτηριστικών ενός θεματικού επιπέδου (layer), κάνουμε δεξί κλικ στο αντίστοιχο επίπεδο του πίνακα περιεχομένων και έπειτα [Open Attribute_Table].

FID	Shape	LU_CODE	sample1	PNTPOLYCNT	AREA	PERIMETER	dokimast
0	Polygon	500	0	0	1567149.59157	10974.853699	0
1	Polygon	700	1	1	13372.205817	1375.324304	0
2	Polygon	400	0	0	16845.38488	847.027897	0
3	Polygon	200	7	7	510486.707256	8444.395366	3
4	Polygon	100	0	0	35642.017801	1094.834372	0
5	Polygon	400	0	0	18571.47012	880.516248	0
6	Polygon	700	0	0	13834.888678	1373.366942	0
7	Polygon	200	4	4	316351.196043	4016.11379	3
8	Polygon	400	0	0	1831.296251	223.597907	0
9	Polygon	400	0	0	6954.370084	473.216148	0
10	Polygon	400	3	3	71453.579147	2892.583888	1
11	Polygon	100	0	0	36578.042203	1211.132348	0
12	Polygon	700	1	1	13033.164825	1208.199666	0
13	Polygon	400	0	0	16051.346736	1031.87589	0
14	Polygon	200	9	9	197739.876443	3189.999727	3
15	Polygon	400	0	0	7710.087967	621.130043	0
16	Polygon	200	3	3	141559.819748	3066.388864	1
17	Polygon	100	2	2	47310.927651	1787.968237	1
18	Polygon	300	0	0	8046.378958	478.656504	0
19	Polygon	400	3	3	174124.945451	4960.32633	0
20	Polygon	400	2	2	43202.402243	1461.476439	0
21	Polygon	400	0	0	31941.688385	1486.064995	0
22	Polygon	300	0	0	6249.732887	323.438433	0
23	Polygon	300	0	0	927.812814	134.86296	0
24	Polygon	400	2	2	141356.240917	3392.423245	1
25	Polygon	100	0	0	4447.275321	298.634375	0
26	Polygon	200	7	7	349215.905775	7082.023817	3

σχήμα 3.13: Μέρος του πίνακα περιγραφικών χαρακτηριστικών του επιπέδου [theme2]

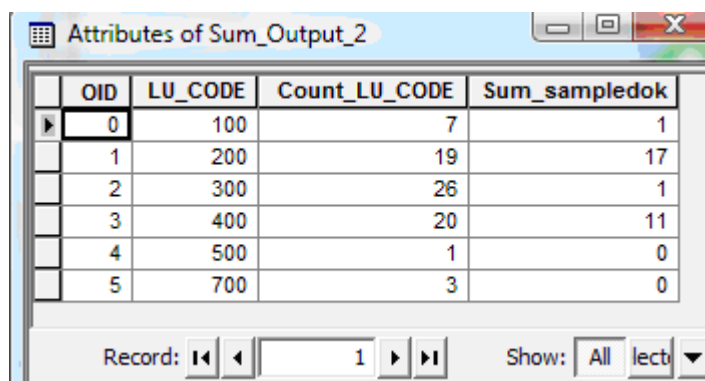
Πλέον αυτό που μένει είναι να γίνει άθροισμα των σημείων της δοκιμαστικής δειγματοληψίας των επιμέρους πολυγώνων που είναι της ίδιας χρήσης, π.χ. σημεία δοκιμαστικής που αντιστοιχούν σε δασικές εκτάσεις, συγκεντρώνοντας έτσι τον συνολικό αριθμό σημείων της δοκιμαστικής δειγματοληψίας ανά χρήση γης.

Στον πίνακα περιγραφικών χαρακτηριστικών του επιπέδου [theme2], μαρκάροντας τη στήλη που δημιουργήθηκε από τη προηγούμενη διαδικασία (dokimastiko), επιλέγεται με δεξί κλικ **summarize**. Το παράθυρο διαλόγου που εμφανίζεται συμπληρώνεται όπως παρακάτω:

σχήμα 3.14: παράθυρο διαλόγου σύνοψης (summarize)

- 1.Πεδίο σύνοψης επιλέγεται η στήλη **[LU_CODE]**, αφού η σύνοψη θα γίνει ως προς τις χρήσεις γης
- 2.Στη στήλη που δημιουργήθηκε κατά τη διαδικασία *count points in polygons* (**dokimastiko**) μαρκάρεται το **[sum]**
- 3.Επιλέγεται το όνομα και η θέση του πίνακα εξόδου.

Απαντώντας θετικά στην εμφάνιση του πίνακα εξόδου, πλέον στην καρτέλα **[source]** του πίνακα περιεχομένων, μπορούμε να ανοίξουμε τον πίνακα που δημιουργήθηκε:



OID	LU_CODE	Count_LU_CODE	Sum_sampledok
0	100	7	1
1	200	19	17
2	300	26	1
3	400	20	11
4	500	1	0
5	700	3	0

σχήμα 3.15: Πίνακας σημείων ανά χρήση δοκιμαστικής δειγματοληψίας

Στήλη LU_CODE (Land Use Code): Σύμφωνα με την κωδικοποίηση που ακολουθήθηκε αντιστοιχούν οι εξής τιμές σε κάθε κατηγορία χρήσης γης:

- 100 για την αρδευόμενη γεωργική γη
- 200 για τις θαμνώδεις εκτάσεις
- 300 για τους βοσκότοπους
- 400 για τις δασικές εκτάσεις
- 500 για την υδάτινη επιφάνεια (δεν παίρνει μέρος στην δειγματοληψία)
- 700 για τις δενδροκαλλιέργειες

Στήλη Count LU_CODE: Ο συνολικός αριθμός πολυγώνων που αντιστοιχούν σε κάθε κατηγορία χρήσης γης.

Στήλη Sum points: Ο συνολικός αριθμός των σημείων της δειγματοληψίας που αντιστοιχούν σε κάθε χρήση γης (LU_CODE).

*** η τελευταία στήλη του πίνακα ονομάζεται **sum points**, όπου *points* όνομα που δώσαμε στη στήλη που δημιουργήσαμε κατά τη διαδικασία "count points in polygons".*

Παρατηρώντας το χάρτη με τα πολύγωνα των 5 κατηγοριών χρήσης γης, δύο είναι οι χρήσεις που κυριαρχούν, συγκεκριμένα οι θαμνώδεις και οι δασικές εκτάσεις (εξαιρώντας φυσικά τις υδάτινες επιφάνειες που δεν αφορούν την δειγματοληψία).

Παράλληλα, οι δενδροκαλλιέργειες καταλαμβάνουν μια ελάχιστη έκταση και είναι πολύ πιθανό σε μια δειγματοληψία τόσο λίγων σημείων να μην εκπροσωπούνται καν. Το ίδιο ισχύει και για την αρδευόμενη γεωργική γη ή ακόμα και για τους

Η αναλογία της έκτασης αυτής επί της συνολικής είναι βέβαια στα όρια του στατιστικού σφάλματος.

3.4 Εύρεση Απαιτούμενου Μεγέθους Δείγματος

Για την εύρεση του απαιτούμενου μεγέθους δείγματος ή πιο ορθά των επιπλέον σημείων που απαιτούνται, διατίθεται λογιστικό φύλλο (αρχείο excel) στο οποίο μεταφέρονται τα αποτελέσματα της δοκιμαστικής δειγματοληψίας – στήλη [sum points]. (σχήμα 3.16)

Αναγράφοντας τα σημεία ανά χρήση γης όπως βρέθηκαν στην δοκιμαστική δειγματοληψία, επιλέγοντας **επίπεδο εμπιστοσύνης** και **περιθώριο λάθους** υπολογίζεται αυτόματα το ελάχιστο απαιτούμενο – επιπλέον της δοκιμαστικής - δείγμα σημείων της οριστικής απλής τυχαίας δειγματοληψίας [n_{new}].

Για πιο ασφαλείς εκτιμήσεις το οριστικό δείγμα μπορεί να υπολογιστεί παρακάμπτοντας τη δοκιμαστική δειγματοληψία, επιλέγοντας απευθείας τιμές στα πεδία του αντίστοιχου φύλλου excel:

- **P = 50** : τιμή που δίνει το μέγιστο αριθμό σημείων για δεδομένο επίπεδο εμπιστοσύνης και περιθώριο σφάλματος.
- **Z = 1.96** :αν επιλέξουμε επίπεδο εμπιστοσύνης 95%

Για την εύρεση του Z, ανάλογα με το επίπεδο εμπιστοσύνης που θα επιλέξουμε χρησιμοποιείται ο πίνακας που παρατίθεται στο παράρτημα.

- **E = 2** :αν το περιθώριο σφάλματος που μας ικανοποιεί είναι 2%

Με αυτές τις επιλογές θα λάβουμε το μέγιστο δείγμα.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Απαιτούμενο μέγεθος δείγματος														
2	*Ο υπολογισμός του απαιτούμενου δείγματος γίνεται μετά την πραγματοποίηση δοκιμαστικής δειγματοληψίας														
3															
4	$n_{new} = \frac{z^2 \cdot \hat{p}_{\%} \cdot \hat{q}_{\%}}{e^2}$														
5															
6															
7															
8															
9	n_{new} : Απαιτούμενο μέγεθος δείγματος για επίπεδο εμπιστοσύνης α , Z και περιθώριο λάθους e (πχ $\alpha = 95\%$, $Z = 1.96$ και $e = 2\%$)														
10	$p\%$: Η καλύτερη εκτίμηση του ποσοστού της κατηγορίας χρήσεων γης, για την οποία, είναι το πλησιέστερο στο 50%, σε σχέση														
11	με τις καλύτερες εκτιμήσεις των ποσοστών των άλλων χρήσεων.														
12	$q\% = 100 - p\%$														
13															
14	Συμπληρώνονται μόνο τα πεδία με κοκκινο χρώμα!														
15															
16	Το αποτέλεσμα της δοκιμαστικής (σημεία ανά χρήση γης) που ως sum_output_** στον πίνακα περιεχομένων τοποθετούμε εδώ για														
17	να βρούμε τα ποσοστά ανα χρήση														
18															
19	Κωδικός χρήσης (LU_CODE)														
20		Σημεία		Ποσοστά											
21	100→	1		3.33 %											
22	200→	17		56.67 %											
23	300→	1		3.33 %											
24	400→	11		36.67 %											
25	700→	0		0.00 %											
26	νολο σημείων δοκιμαστικ	30													
27															
28	$p\% =$	56.67		Η τιμή του $p\%$ είναι αυτή της δοκιμαστικής δειγματοληψίας που έχει προηγηθεί (αναγράφουμε το ποσοστό που είναι πλησιέστερο στο 50%)											
29	$z =$	1.96													
30	$e\% =$	2		$n_{new} =$	2328	επιπλέον σημεία									
31															

σχήμα 3.16: Φύλλο Excel για τον αυτόματο υπολογισμό απαιτούμενου μεγέθους δείγματος (σημεία δοκιμαστικής + επιπλέον σημεία)

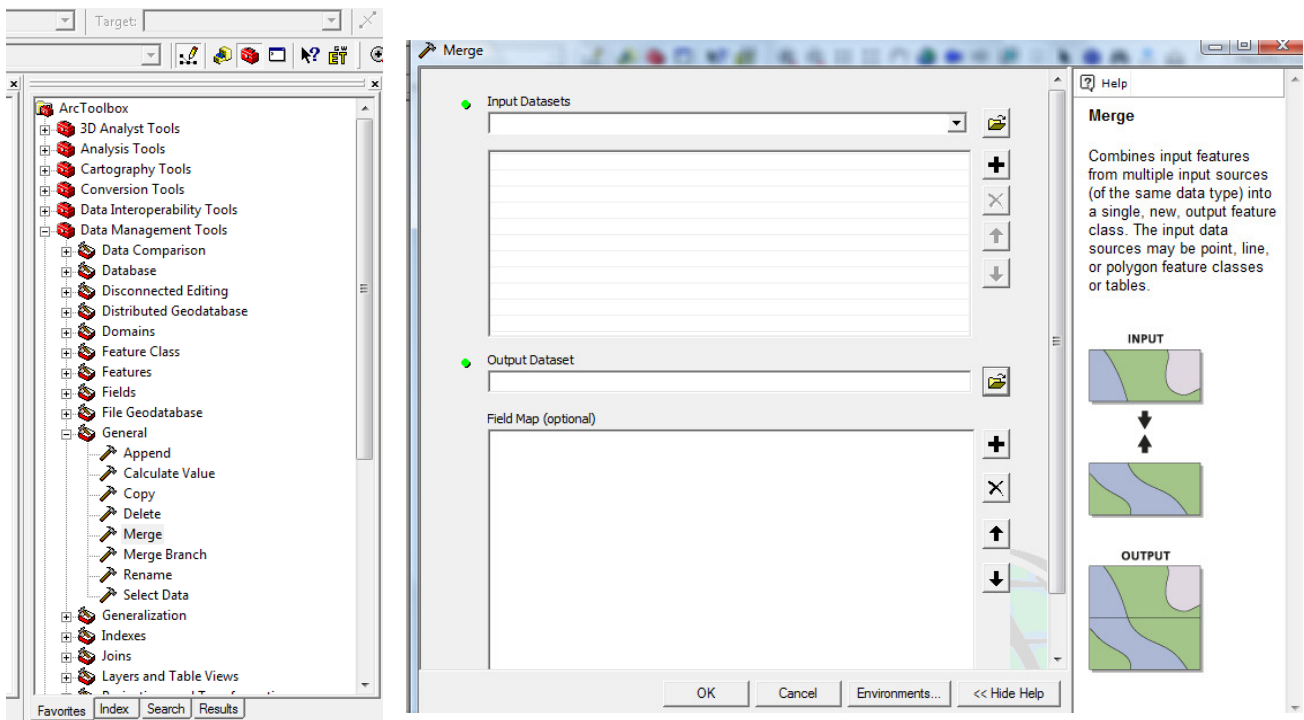
3.5 Εφαρμογή Απλής Τυχαίας Δειγματοληψίας

Μετά την πραγματοποίηση της δοκιμαστικής δειγματοληψίας και την εύρεση του ελάχιστου απαιτούμενου μεγέθους δείγματος που ικανοποιεί τις απαιτήσεις μας, μπορεί να πραγματοποιηθεί η οριστική δειγματοληψία. Η διαδικασία εφαρμογής της περιλαμβάνει τα εξής:

- Από το φύλλο excel υπολογίζουμε τα επιπλέον σημεία δειγματοληψίας που απαιτούνται.
- Πραγματοποιείται τυχαία δειγματοληψία με N_{new} αριθμό σημείων (επιπλέον σημεία, όπως υπολογίστηκαν).
- Τα δύο πλέον σημειακά επίπεδα που υπάρχουν στον πίνακα περιεχομένων συγχωνεύονται σε ένα, που αποτελεί το δείγμα της οριστικής απλής τυχαίας δειγματοληψίας.

- Η συγχώνευση των δύο σημειακών επιπέδων γίνεται με την εντολή **merge**, από το arctoolbox.

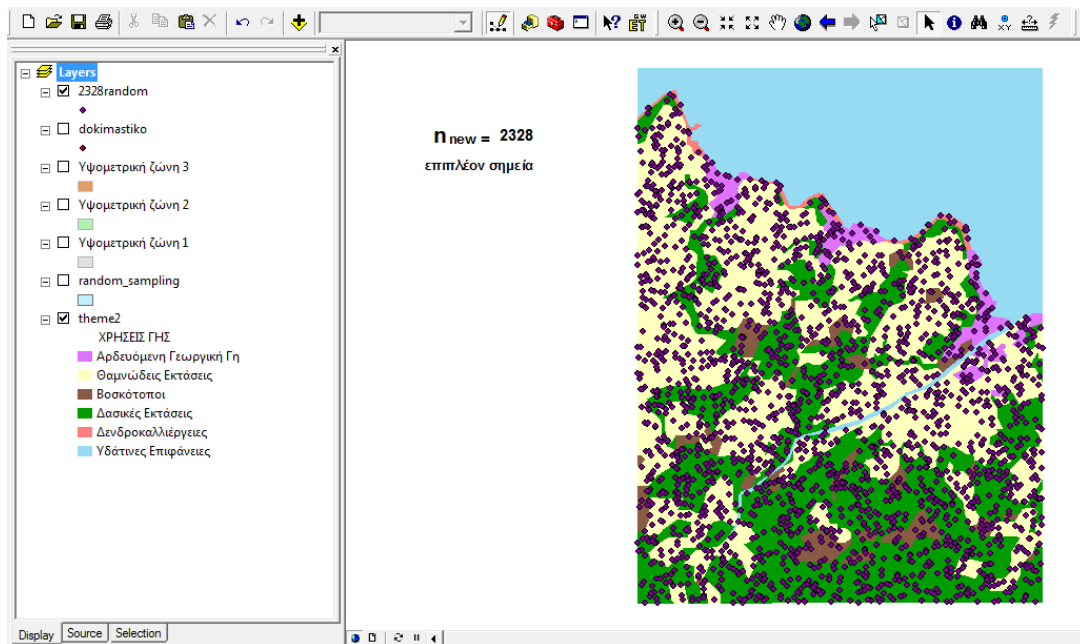
[Data management tools > general > merge].



σχήμα 3.17: παράθυρο διαλόγου merge

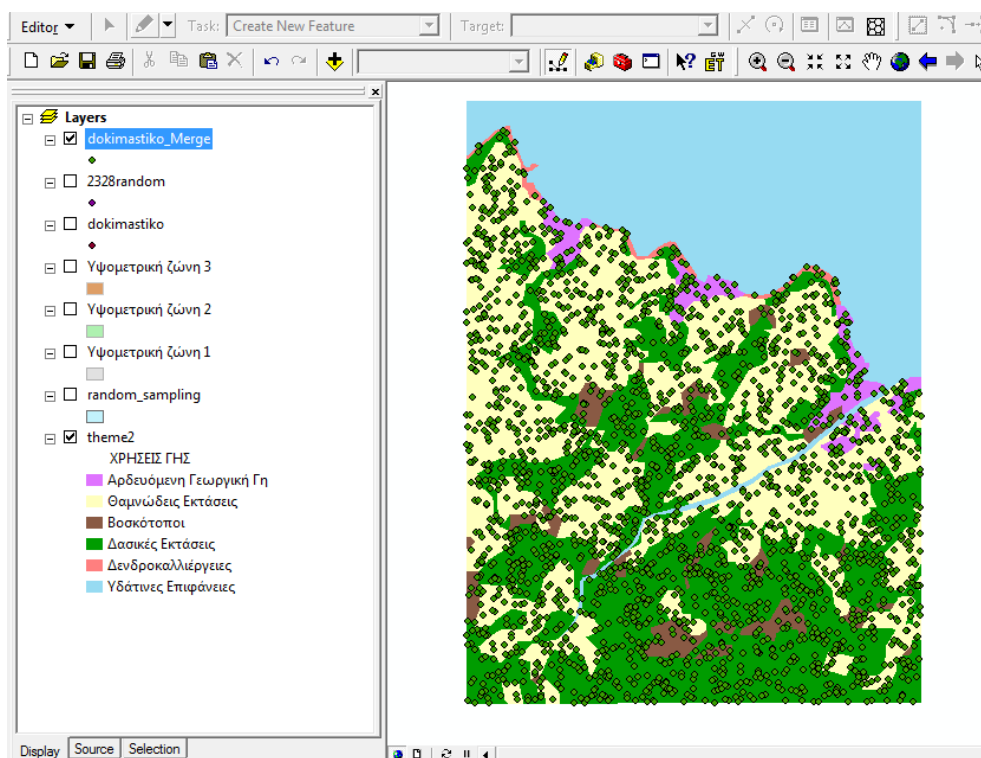
Στο παράθυρο διαλόγου του merge επιλέγονται τα δύο σημειακά επίπεδα.

Η διαδικασία συνεχίζεται κατά τα γνωστά στο νέο επίπεδο που δημιουργήθηκε από τη συγχώνευση. (count points in polygons, summarize).



σχήμα 3.18: Τα απαιτούμενα επιπλέον σημεία

Με την εντολή merge δημιουργείται το νέο σημειακό επίπεδο επίπεδο που συγχωνεύει δοκιμαστικό και επιπλέον δείγμα.

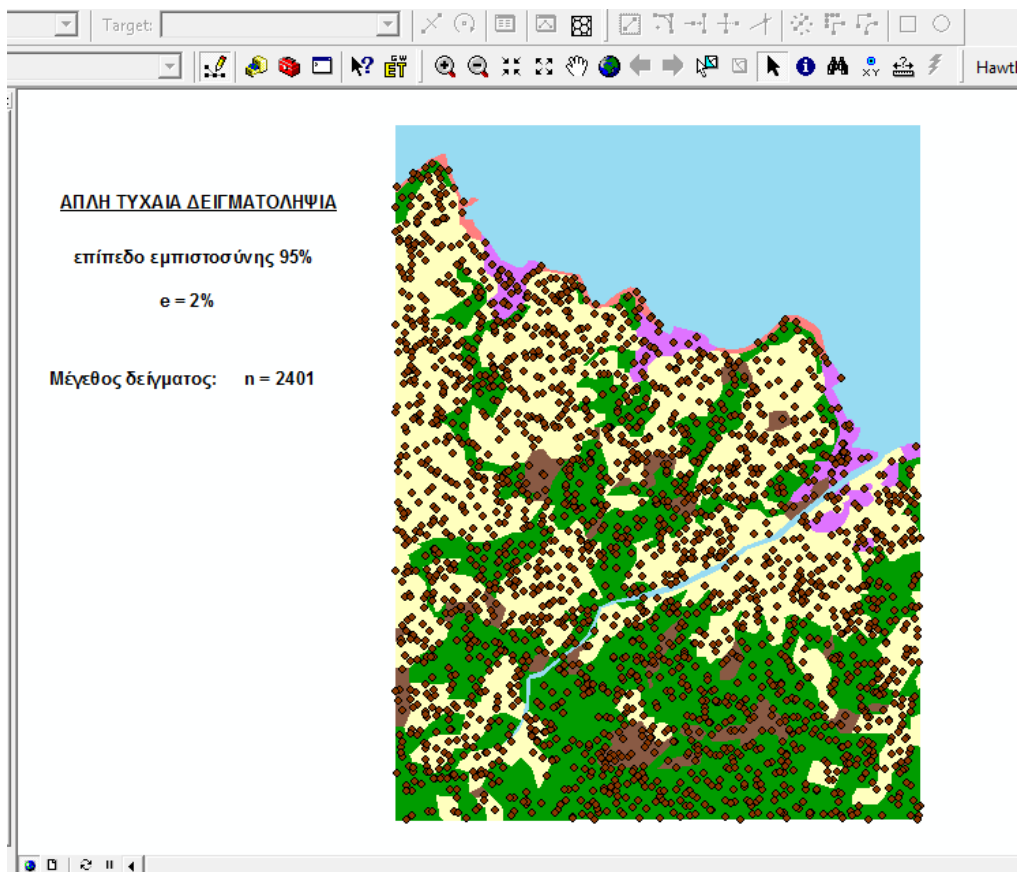


σχήμα 3.19: Οριστικό δείγμα 2358 σημείων

Η απόκλιση στα ποσοστά των διαφόρων κατηγοριών κατά τη πραγματοποίηση δύο ή περισσότερων δοκιμαστικών δειγματοληψιών 30 σημείων μπορεί να είναι αρκετά μεγάλη.

Για να έχουμε τη μέγιστη ασφάλεια, με τμήμα τα περισσότερα σημεία δείγματος, μπορεί να πραγματοποιηθεί απευθείας οριστική δειγματοληψία με $n_{max} = \dots$ σημεία (στη περίπτωση αυτή λαμβάνεται $P\% = 50$).

19	Κωδικός χρήσης (LU_CODE)								
20		Σημεία	Ποσοστά						
21	100→	1	3.33 %						
22	200→	17	56.67 %						
23	300→	1	3.33 %						
24	400→	11	36.67 %						
25	700→	0	0.00 %						
26	συνολο σημείων δοκιμαστικ	30							
27									
28	p%=	56.67		Η τιμή του p% είναι αυτή της δοκιμαστικής δειγματοληψίας που έχει προηγηθεί (αναγράφουμε το ποσοστό που είναι πλησιέστερο στο 50%)					
29	z=	1.96							
30	e%=	2		n_{new} =	2328	επιλέγον σημεία			
31									
32	Για μεγαλύτερη ασφάλεια χρησιμοποιείται στην οριστική δειγματοληψία			n_{max} =	2401				
33						περιθώριο λαθους			
34						E=	1.98 %		
35									
36									



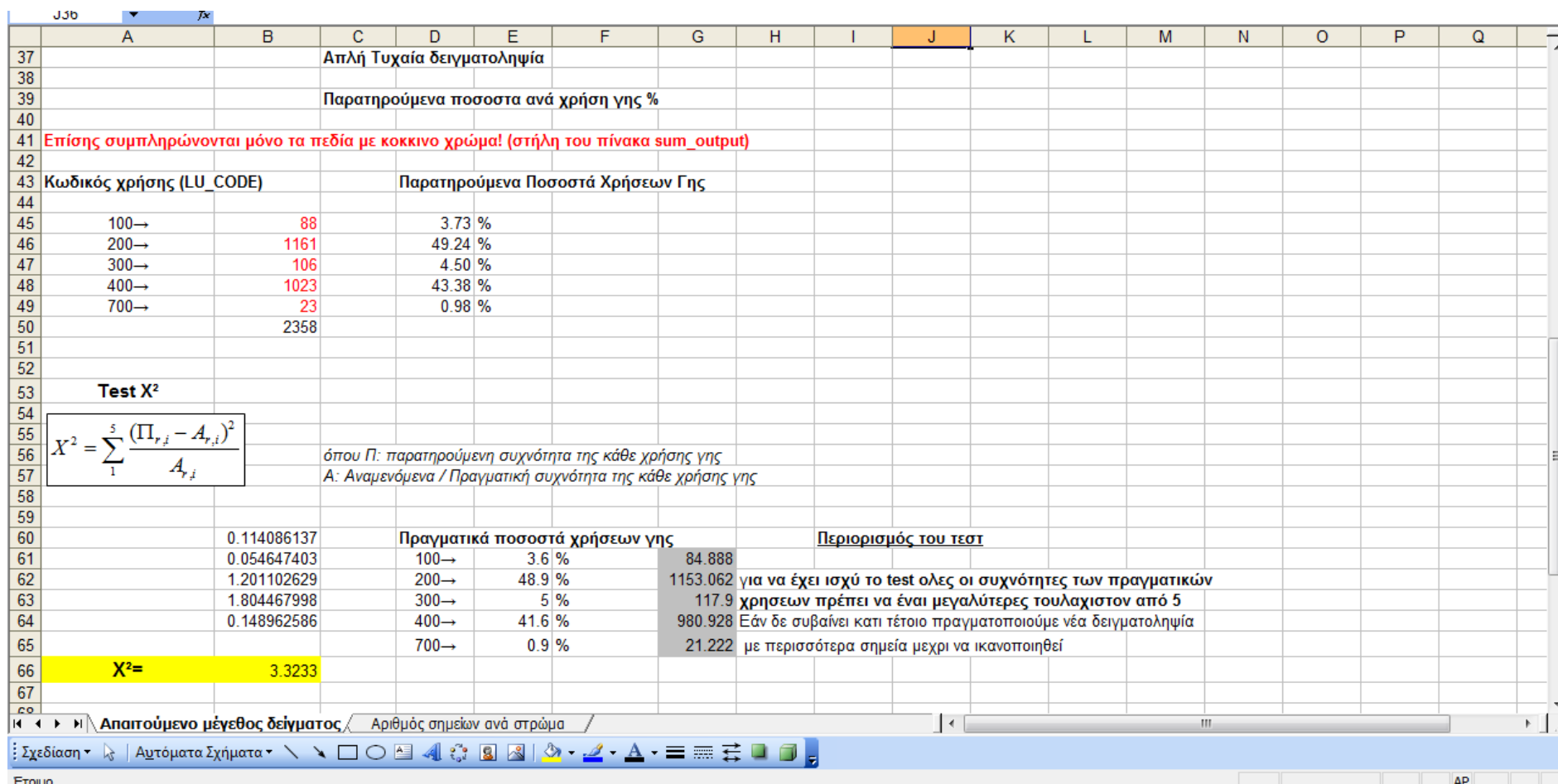
σχήμα 3.20: Απλή τυχαία δειγματοληψία 2401 σημείων

OID	LU_CODE	Count_LU_CODE	Sum_sample100
0	100	7	88
1	200	19	1161
2	300	26	106
3	400	20	1023
4	500	1	0
5	700	3	23

σχήμα 3.21: σημεία οριστικού δείγματος ανά κατηγορία χρήσης

Τα αποτελέσματα της Απλής Τυχαίας Δειγματοληψίας φαίνονται στον παραπάνω πίνακα (σχήμα 3.21) δηλαδή η αντιστοιχία χρήσεων γης (LU_CODE) και σημείων της τυχαίας δειγματοληψίας (Sum_sample100). Για την ολοκλήρωση της εφαρμογής αντιγράφουμε τα παραπάνω αποτελέσματα στη κατάλληλη θέση του υπολογιστικού φύλλου που διατίθεται και έτσι υπολογίζονται τα παρατηρούμενα ποσοστά της κάθε χρήσης.

Παράλληλα υπολογίζεται αυτόματα η τιμή του test χ^2 για τη παραπάνω δειγματοληψία. Συγκρίνοντας τη τιμή του τεστ με τον πίνακα που διατίθεται στο παράρτημα και περιλαμβάνει τις κρίσιμες τιμές του χ^2 για δεδομένο επίπεδο σημαντικότητας (α) και βαθμούς ελευθερίας d , στη περίπτωση της δειγματοληψίας είναι $d = 5 - 1 = 4$, αφού έχουμε 5 διαφορετικές κατηγορίες χρήσεων γης (σχήμα 3.22).



σχήμα 3.22: Υπολογισμός παρατηρούμενων ποσοστών χρήσεων γης και Test X²

Στο φύλλο excel υπολογίζεται και το διάστημα εμπιστοσύνης για κάθε κατηγορία χρήσης, δηλαδή το διάστημα μέσα στο οποίο βρίσκεται η πραγματική τιμή του ποσοστού κάθε κατηγορίας με πιθανότητα ίση με το επίπεδο εμπιστοσύνης που έχουμε θέσει.

68										
69										
70	$\hat{p} \pm z_p \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$		ΔΙΑΣΤΗΜΑ ΕΝΤΟΣ ΤΟΥ ΟΠΟΙΟΥ ΒΡΙΣΚΕΤΑΙ ΤΟ ΠΡΑΓΜΑΤΙΚΟ ΠΟΣΟΣΤΟ (%) ΚΑΘΕ ΚΑΤΗΓΟΡΙΑΣ ΓΙΑ ΜΕΓΕΘΟΣ ΔΕΙΓΜΑΤΟΣ n ΚΑΙ ΓΙΑ ΕΠΙΠΕΔΟ ΕΜΠΙΣΤΟΣΥΝΗΣ ΠΟΥ ΕΧΕΙ ΟΡΙΣΤΕΙ							
71										
72										
73										
74										
75										
76										
77	$\hat{p}_{100} =$	0.036651395	$\hat{p}_{200} =$	0.483549	$\hat{p}_{300} =$	0.044148	$\hat{p}_{400} =$	0.426072	$\hat{p}_{700} =$	0.009579
78	$1 - \hat{p}_{100} =$	0.963348605	$1 - \hat{p}_{200} =$	0.516451	$1 - \hat{p}_{300} =$	0.955852	$1 - \hat{p}_{400} =$	0.573928	$1 - \hat{p}_{700} =$	0.990421
79	διάστημα \pm	0.75	διάστημα \pm	2.00	διάστημα \pm	0.82	διάστημα \pm	1.98	διάστημα \pm	0.39
80										
81										
82										
83										

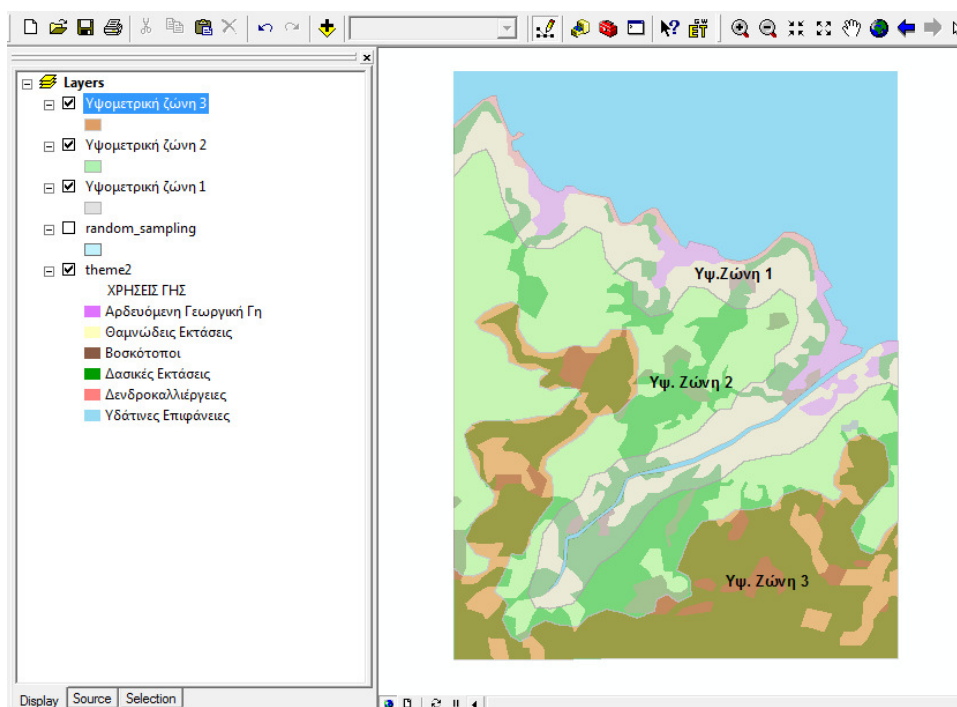
σχήμα 3.23: υπολογισμός διαστήματος εμπιστοσύνης

3.6 Στρωματοποιημένη Τυχαία Δειγματοληψία

Στη περίπτωση αυτή η δειγματοληψία επιτυγχάνεται με τη στρωματοποίηση του υποβάθρου σε υποπεριοχές και στη συνέχεια σε κάθε μία από αυτές γίνεται τυχαία επιλογή δειγμάτων (απλή τυχαία δειγματοληψία) με τον γνωστό τρόπο.

Το υπόβαθρο στρωματοποιήθηκε σε 3 υποπεριοχές που αντιστοιχούν στις τρεις υψομετρικές ζώνες του χάρτη (σχήμα 3.24) :

- Υψομετρική ζώνη 1
- Υψομετρική ζώνη 2
- Υψομετρική ζώνη 3



σχήμα 3.24: Δειγματοληπτικό Υπόβαθρο Στρωματοποιημένης Τυχαίας Δειγματοληψίας

Το εμβαδό και η περίμετρος των υποπεριοχών (σε m² και m αντίστοιχα) βρίσκονται στους πίνακες περιγραφικών χαρακτηριστικών των επιπέδων.

Το συνολικό μέγεθος δείγματος της στρωματοποιημένης δειγματοληψίας είναι το ίδιο με αυτό που χρησιμοποιήθηκε και στην απλή τυχαία. [n_{max}].

Σε περίπτωση που απαιτείται διαφορετικό επίπεδο εμπιστοσύνης ή περιθώριο λάθους θα πρέπει να προσδιοριστούν οι τιμές στο φύλλο excel. Στη δεύτερη καρτέλα του φύλλου excel υπολογίζεται αυτόματα ο αριθμός των τυχαίων σημείων που θα ληφθούν ανά στρώμα. Ο αριθμός των σημείων είναι ανάλογος της επιφάνειας του κάθε στρώματος - υψομετρικής ζώνης (φυσική στρωματοποίηση).

Για την Υψομετρική ζώνη 0 – 100 μ. είναι **n1**
 Για την Υψομετρική ζώνη 100 – 200 μ. είναι **n2**
 Για την Υψομετρική ζώνη 200 – 300 μ. είναι **n3**

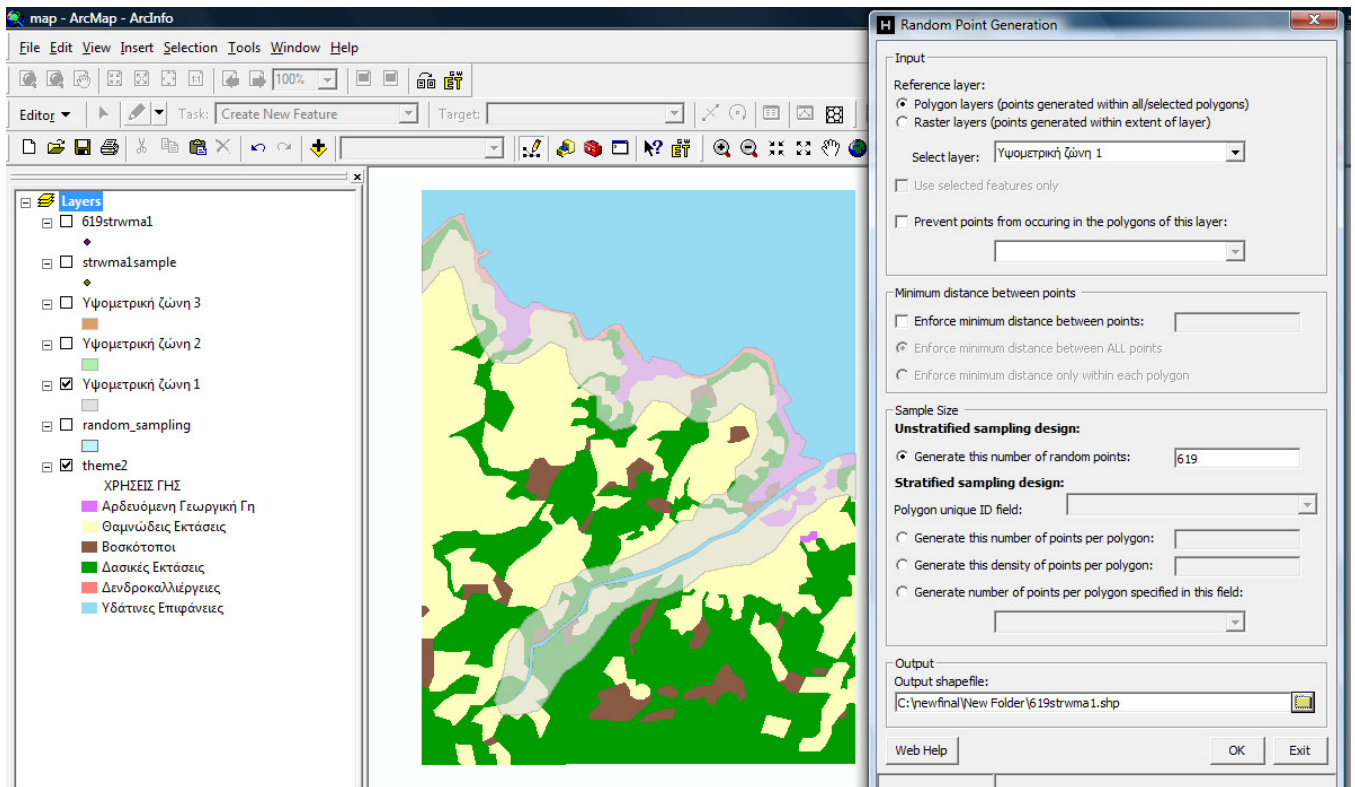
Έτσι πλέον μένει να πραγματοποιηθούν **3 απλές τυχαίες δειγματοληψίες** σε κάθε υψομετρική ζώνη.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Στρωματοποιημένη δειγματοληψία												
2	Αριθμός σημείων ανά στρώμα (ανάλογος της επιφάνειας του κάθε στρώματος)												
3													
4	Αφού γίνει η δοκιμαστική δειγματοληψία (μια φορά για τις δύο μεθόδους) υπολογίζεται ο αριθμός σημείων ανά στρώμα.												
5	Στη στρωματοποιημένη θα γίνουν τρεις απλές τυχαίες δειγματοληψίες στα ισάριθμα στρώματα, με αριθμό σημείων στη κάθε μία που υπολογίζεται εδώ												
6													
7	$n_1 = n_{new} \cdot \text{Εμβ}1 / \Sigma\text{Ε}$												
8	$n_2 = n_{new} \cdot \text{Εμβ}2 / \Sigma\text{Ε}$												
9	$n_3 = n_{new} \cdot \text{Εμβ}3 / \Sigma\text{Ε}$												
10	όπου n1, n2, n3 ο αριθμός των σημείων σε κάθε στρώμα nnew: το απαιτούμενο μέγεθος δείγματος που βρέθηκε στα προηγούμενα Εμβ1, Εμβ2, Εμβ3: Τα εμβαδά των στρωμάτων ΣΕ: το συνολικό εμβαδό του δειγματοληπτικού υποβάθρου												
11	Εμβ1= 1099680 0.258 Εμβ2= 1721448 0.404 Εμβ3= 1443004 0.338 ΣΕ= 4264132 1.000												
12	Αριθμός σημείων δειγματοληψίας για κάθε ζώνη												
13	n1=	619											
14	n2=	969											
15	n3=	813											
16		2401											
17													
18													
19													
20													

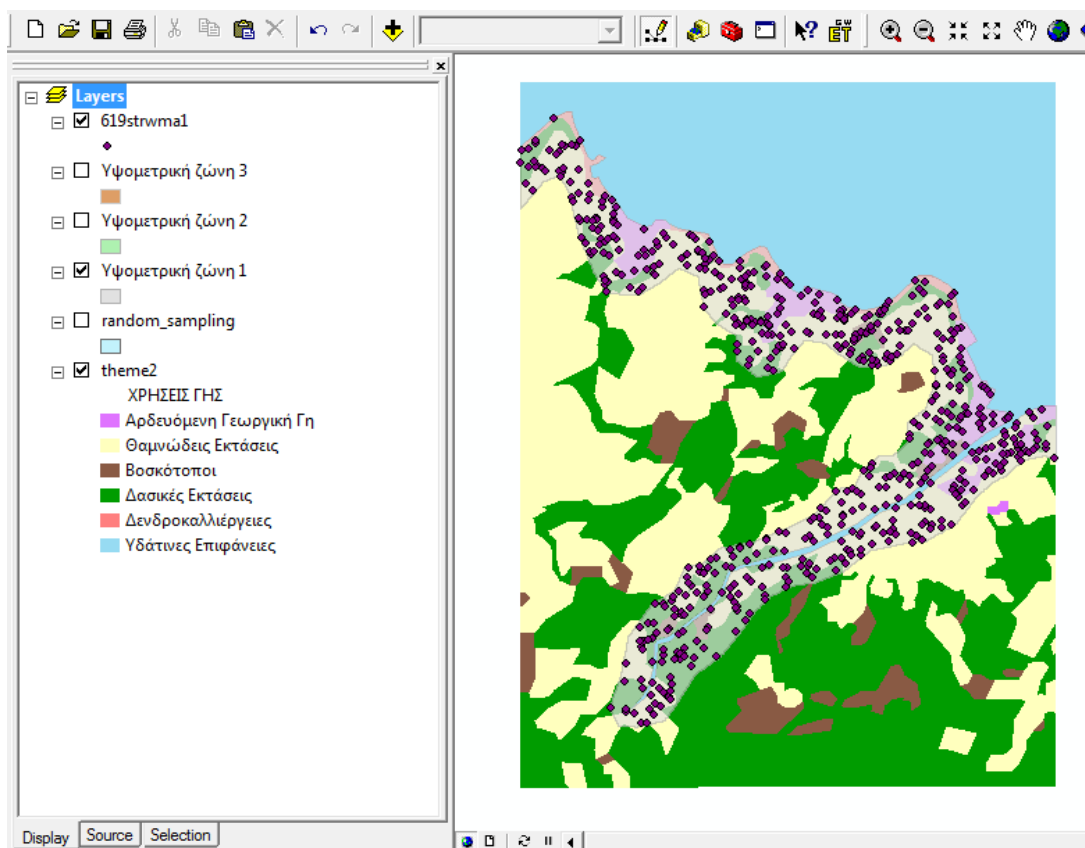
σχήμα 3.25: Απόσπασμα του φύλλου excel για τον καταμερισμό των σημείων ανά στρώμα

- Υψομετρική ζώνη 1 (1^ο στρώμα)

Ο αριθμός των τυχαίων σημείων που λαμβάνονται στο πρώτο στρώμα είναι **n1**.



Κατά αντιστοιχία με την απλή τυχαία, από την μπάρα εργαλείων Hawth's Tools επιλέγεται η γεννήτρια τυχαίων σημείων. Το επίπεδο που επιλέγεται στο παράθυρο διαλόγου του random point generator [select layer] είναι η υψομετρική ζώνη 1 και ο αριθμός τυχαίων σημείων στο παράδειγμα μας 619, όπως υπολογίστηκε αυτόματα στη δεύτερη καρτέλα του φύλλου excel. Ακολουθεί η διαδικασία Count points in polygons όπως περιγράφηκε στην απλή τυχαία και το summarize στη στήλη του πίνακα που δημιουργείται. Αποτέλεσμα ένας πίνακας sum_output όπως στα προηγούμενα.



σχήμα 3.26: Κατανομή τυχαίων σημείων στο πρώτο στρώμα

OID	LU_CODE	Count_LU_CODE	Sum_strwma1
0	100	7	81
1	200	19	343
2	300	26	14
3	400	20	160
4	500	1	0
5	700	3	21

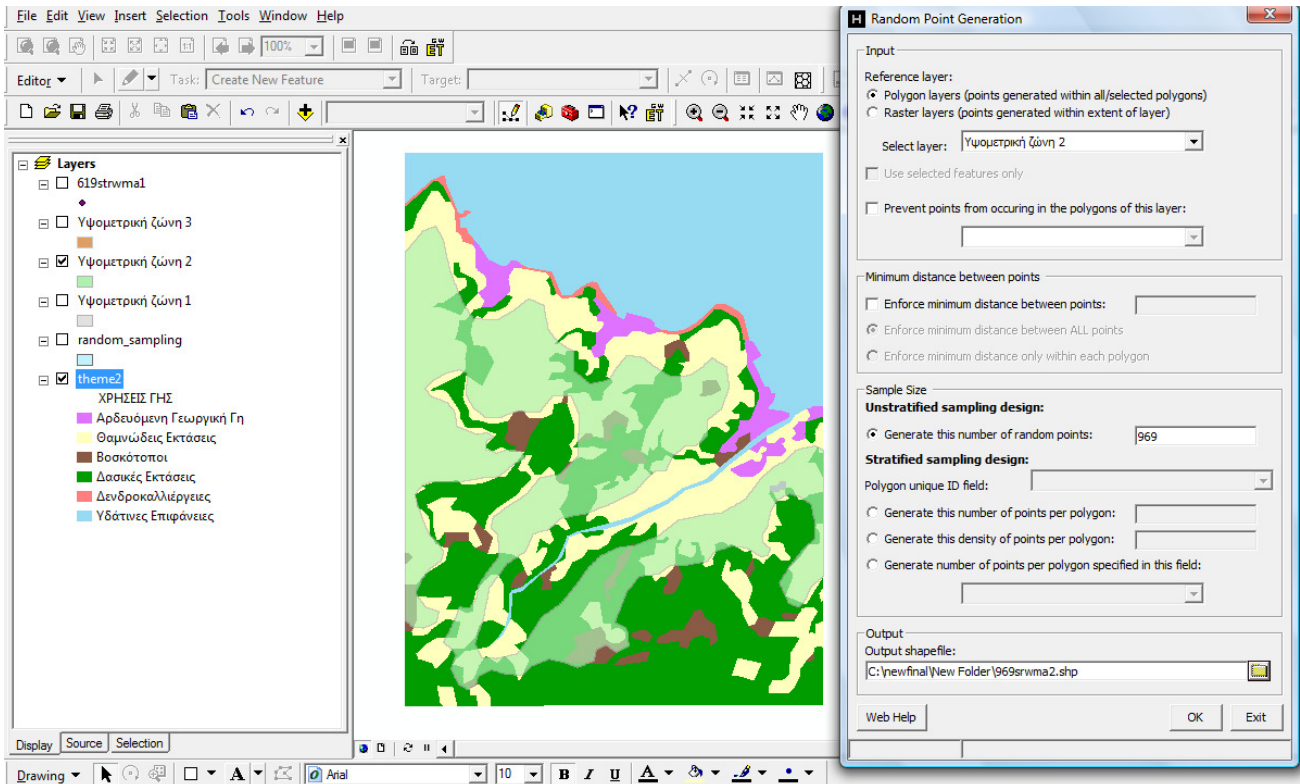
σχήμα 3.27: πίνακας εξόδου πρώτου στρώματος

- Υψομετρική ζώνη 2 (2^ο στρώμα)

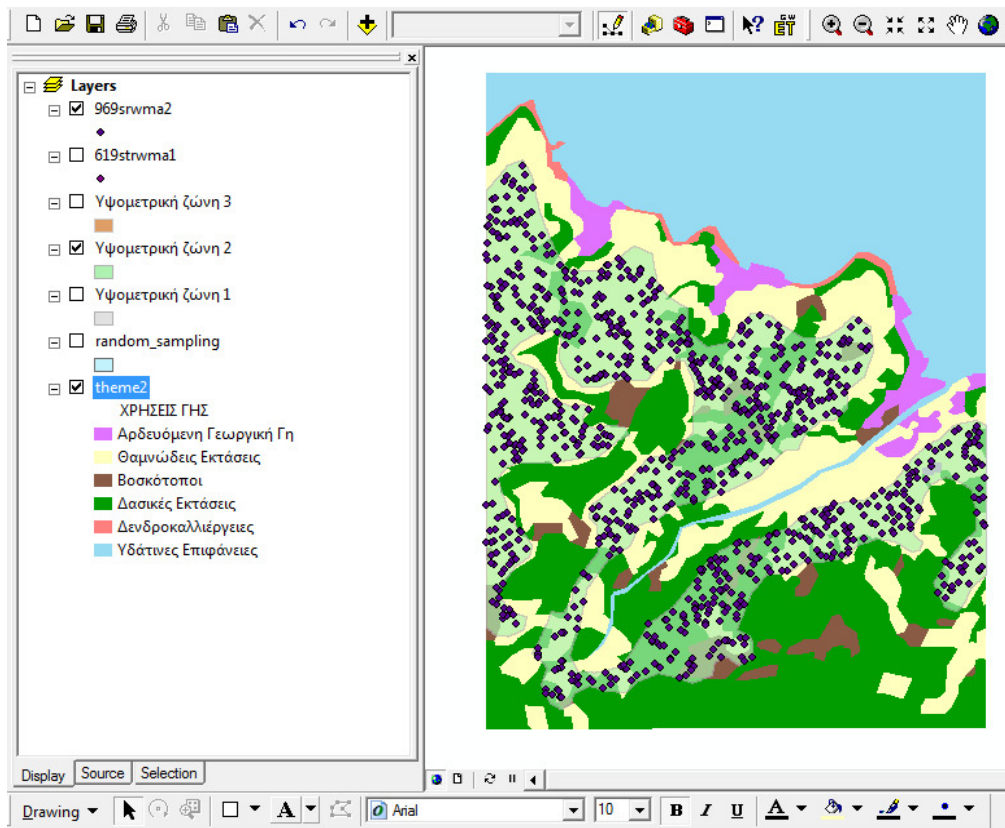
Η διαδικασία δημιουργίας σημείων είναι ίδια, αλλάζοντας το επίπεδο που θα τοποθετηθούν τα τυχαία σημεία (Υψομετρική ζώνη 2).

Ο αριθμός των τυχαίων σημείων που λαμβάνονται στο δεύτερο στρώμα είναι **n2**.

Σύμφωνα με το φύλλο excel στο παράδειγμα μας είναι 969 τυχαία σημεία.



σχήμα 3.28: 2^ο στρώμα δειγματοληψίας



σχήμα 3.29: τυχαία σημεία στο 2^ο στρώμα

Attributes of Sum_Output_6

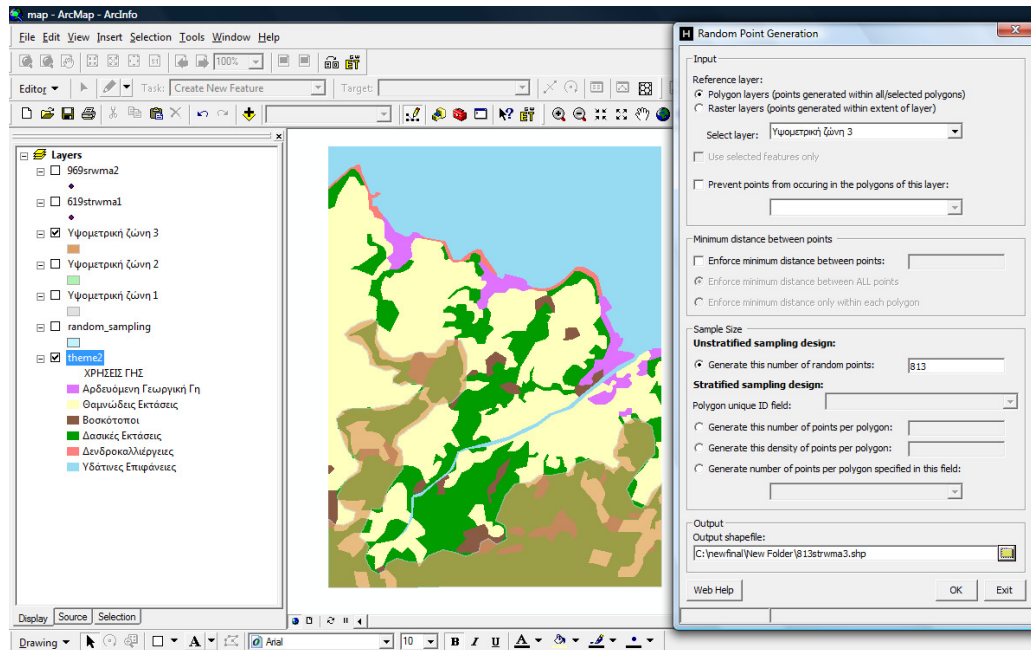
OID	LU_CODE	Count_LU_CODE	Sum_969points
0	100	7	0
1	200	19	685
2	300	26	37
3	400	20	247
4	500	1	0
5	700	3	0

Record: 1 Show: All

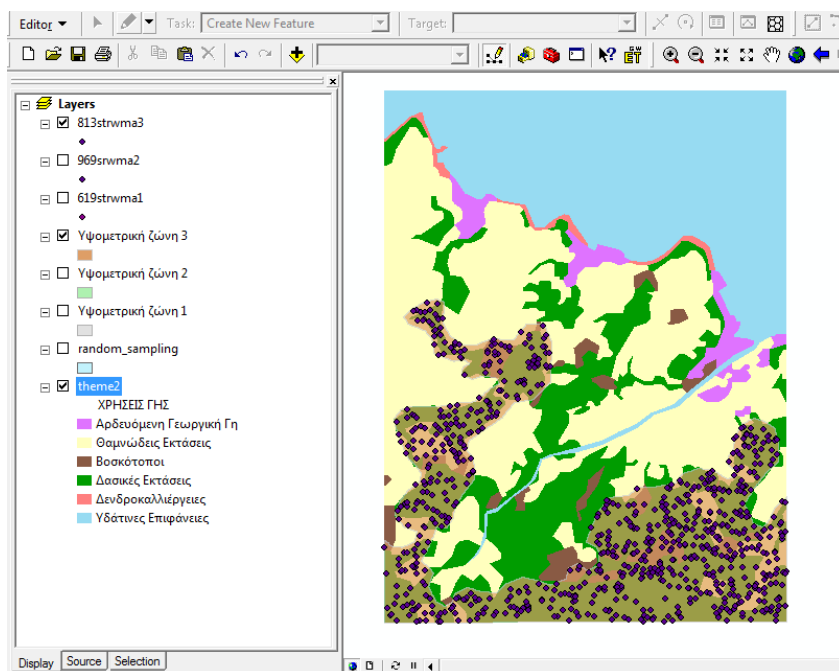
σχήμα 3.30: Πίνακας εξόδου δεύτερου στρώματος

- Υψομετρική ζώνη 3 (3^ο στρώμα)

Ο αριθμός των τυχαίων σημείων που λαμβάνονται στο τρίτο στρώμα είναι **n3**. Συμφωνά με το φύλλο excel στο τρίτο στρώμα αντιστοιχούν 813 σημεία (στο παράδειγμα). Αντίστοιχα επιλέγουμε υπόβαθρο για τη τοποθέτηση των τυχαίων σημείων την τρίτη υψομετρική ζώνη.



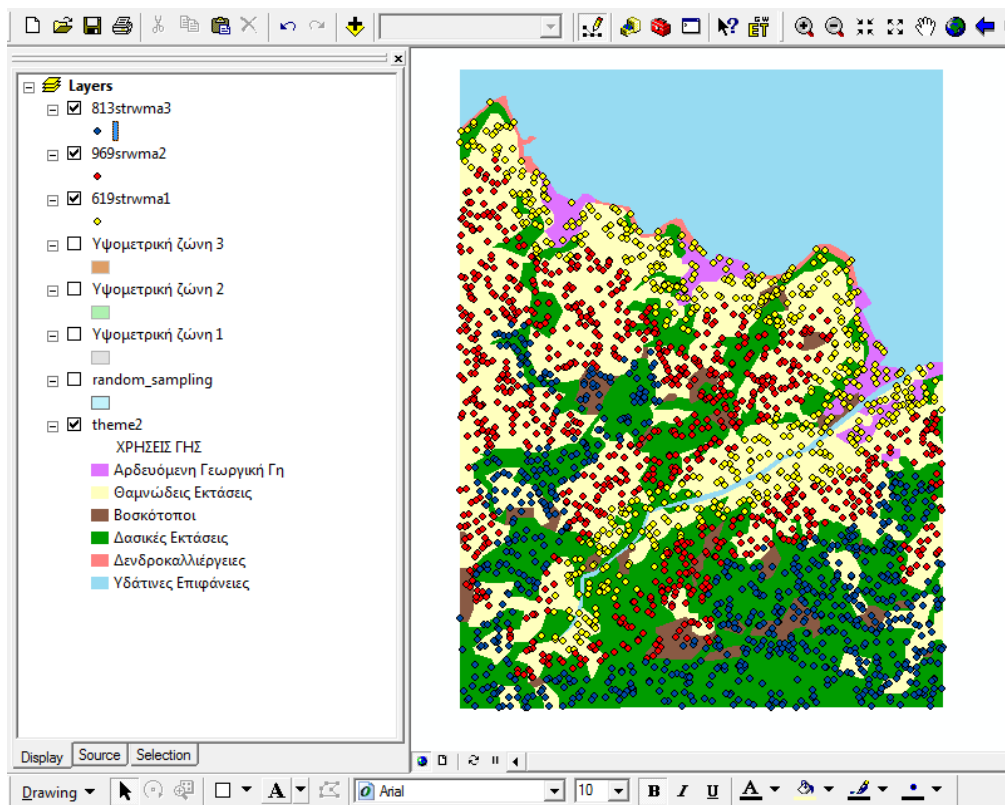
σχήμα 3.31: 3^ο στρώμα δειγματοληψίας



σχήμα 3.32: κατανομή τυχαίων σημείων 3^{ου} στρώματος

OID	LU_CODE	Count_LU_CODE	Sum_813points
0	100	7	0
1	200	19	167
2	300	26	65
3	400	20	581
4	500	1	0
5	700	3	0

σχήμα 3.32: Πίνακας κατανομής σημείων ανά χρήση 3^{ου} στρώματος (πίνακας εξόδου)



σχήμα 3.33: Σημεία Στρωματοποιημένης Τυχαίας Δειγματοληψίας μετά την ολοκλήρωση της διαδικασίας

Τα αποτελέσματα των τριών πινάκων δηλαδή τα σημεία των δειγματοληψιών που αντιστοιχούν σε κάθε χρήση γης, για κάθε υποπεριοχή του δειγματοληπτικού υπόβαθρου **αντιγράφονται** στα αντίστοιχα κελιά του φύλλου excel (αριθμοί με κόκκινο χρώμα). Αυτόματα υπολογίζεται το άθροισμα των σημείων των τριών υποπεριοχών για κάθε χρήση γης (LU_CODE) και τα υπολογίζονται επίσης τα παρατηρούμενα ποσοστά ανά χρήση (σχήμα 3.34).

	A	B	C	D	E	F	G	H	I
31									
32	Στρωματοποιημένη δειγματοληψία								
33									
34	**Αντιγράφουμε από τους πίνακες εξόδου του arcstar τα αποτελέσματα στα αντίστοιχα κελιά με τις κόκκινες τιμές!								
35	(Τα υπόλοιπα υπολογίζονται αυτόματα)								
36	Αριθμός σημείων	1° στρώμα	2° στρώμα	3° στρώμα	Άθροισμα				
37	100→	81	0	0	81				
38	200→	343	685	167	1195				
39	300→	14	37	65	116				
40	400→	160	247	581	988				
41	700→	21	0	0	21				
42		619	969	813	2401				
43									
44									
45	Παρατηρούμενα ποσοστά ανά χρήση γης %								
46									
47	Χρήσεις γης (LU_CODE)								
48	100→	3.37 %							
49	200→	49.77 %							
50	300→	4.83 %							
51	400→	41.15 %							
52	700→	0.87 %							
53									

σχήμα 3.34: απόσπασμα φύλλου excel για τον υπολογισμό παρατηρούμενων ποσοστών ανά χρήση γης της στρωματοποιημένης δειγματοληψίας.

3.7 Συστηματική Δειγματοληψία

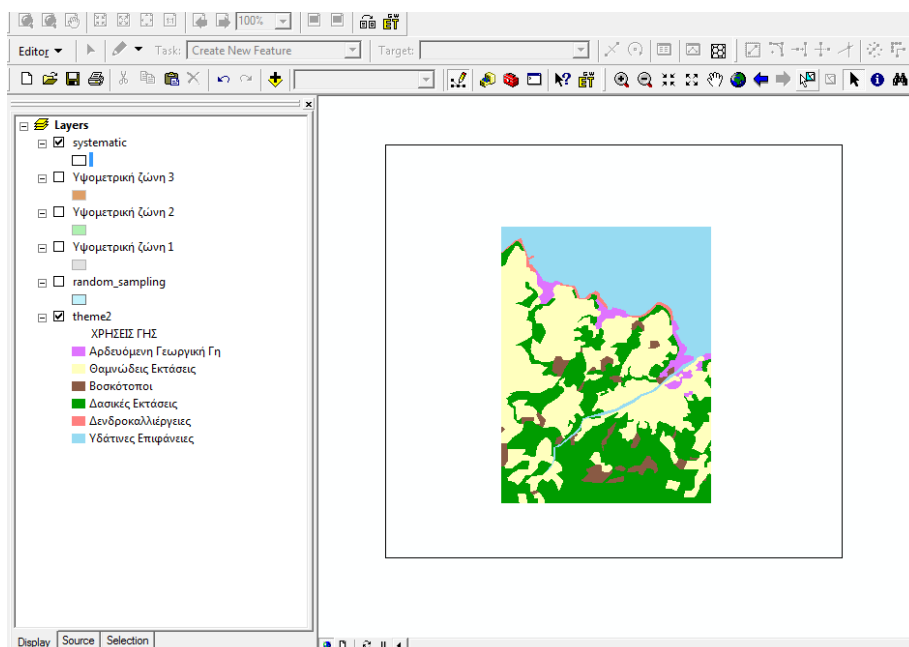
Χρησιμοποιώντας το εργαλείο της μπάρας Hawth [generate regular points], δίνεται η δυνατότητα δημιουργίας σημείων που θα κατανέμονται με ομοιόμορφο τρόπο στο δειγματοληπτικό υπόβαθρο.

Αυτό που μπορούμε να ορίσουμε με τη βοήθεια του εργαλείου είναι η απόσταση – το διάκενο στο σύνολο των σημείων κατά τις δύο διευθύνσεις (κατά X και Y).

Για να μετατρέψουμε τον συστηματικό αυτό τρόπο σε τυχαίο συστηματικό απαιτείται αρχικά να δημιουργήσουμε (edit) ένα τυχαίο ορθογώνιο που να περιβάλλει τον χάρτη και πάνω στο οποίο θα τοποθετηθούν τα σημεία (σχήμα 3.35). Ο λόγος που κάνουμε κάτι τέτοιο είναι ότι το εργαλείο της μπάρας τοποθετεί τα σημεία ξεκινώντας από το πάνω αριστερό άκρο του επιπέδου που του υποδεικνύουμε, έτσι είναι μεν συστηματικά κατανεμημένα τα σημεία αλλά η διαδικασία δεν είναι τυχαία (επανάληψη της διαδικασίας με το ίδιο point spacing τοποθετεί με τον ίδιο ακριβώς τρόπο τα σημεία στο πλαίσιο), κάτι το οποίο δεν είναι επιθυμητό. Το ορθογώνιο θα δημιουργηθεί στο επίπεδο systematic.

Η διαδικασία έχει ως εξής:

- Ενεργοποίηση των εργαλειοθηκών editor και advanced editing
- Εκκίνηση της διαδικασίας [start editor]
- Επιλογή ως στόχου (target) το επίπεδο systematic
- Από την εργαλειοθήκη advanced tools επιλέγεται το rectangle tool, και κατασκευάζεται τυχαίο ορθογώνιο.
- Αποθηκεύεται η εγγραφή (save edits)

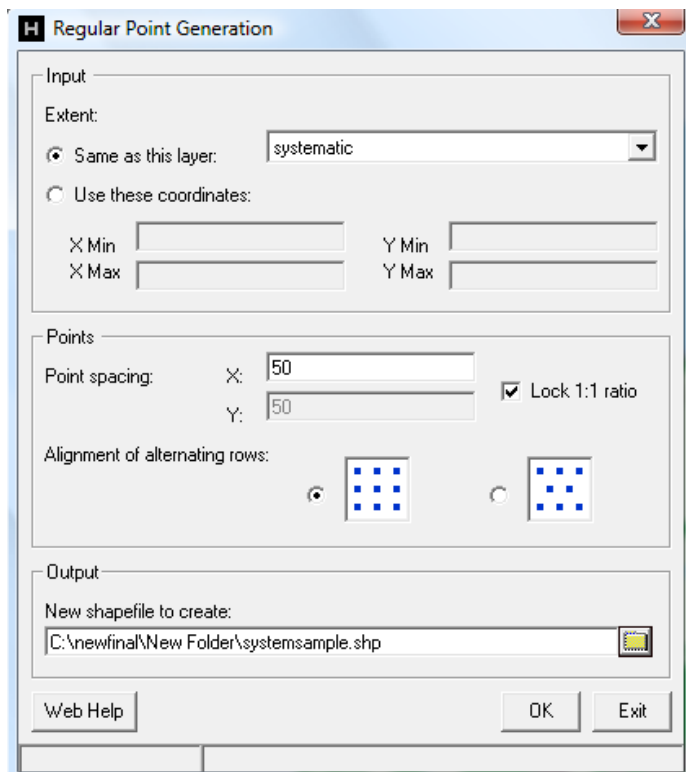


σχήμα 3.35: τυχαίο πλαίσιο

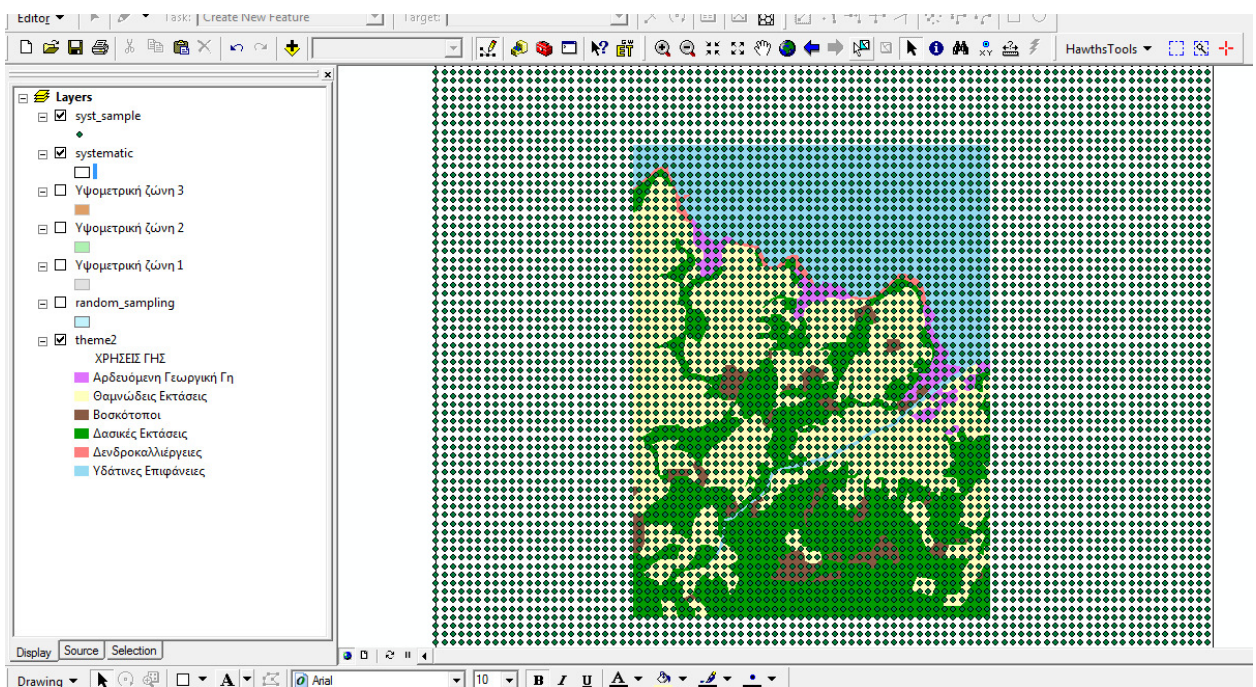
Πλέον στο πίνακα περιγραφικών χαρακτηριστικών του layer systematic υπάρχει το ορθογώνιο.

- Δημιουργία ομοιόμορφα κατανεμημένων σημείων στα όρια του επιπέδου systematic [hawth tools > generate systematic points].

- Επιλογή χωρικού διάκενου κατά X και κατά Y.

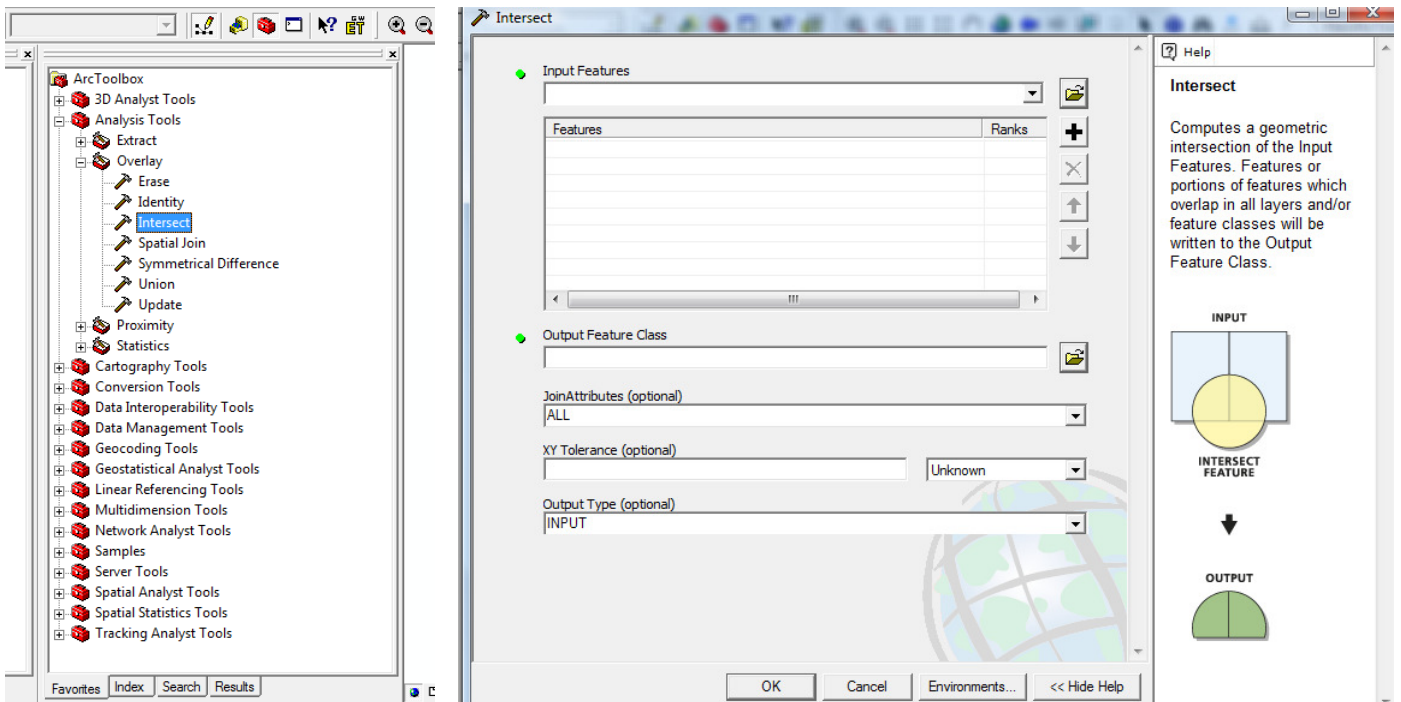


σχήμα 3.36: Επιλογή ορίων κατανομής, διαστήματος



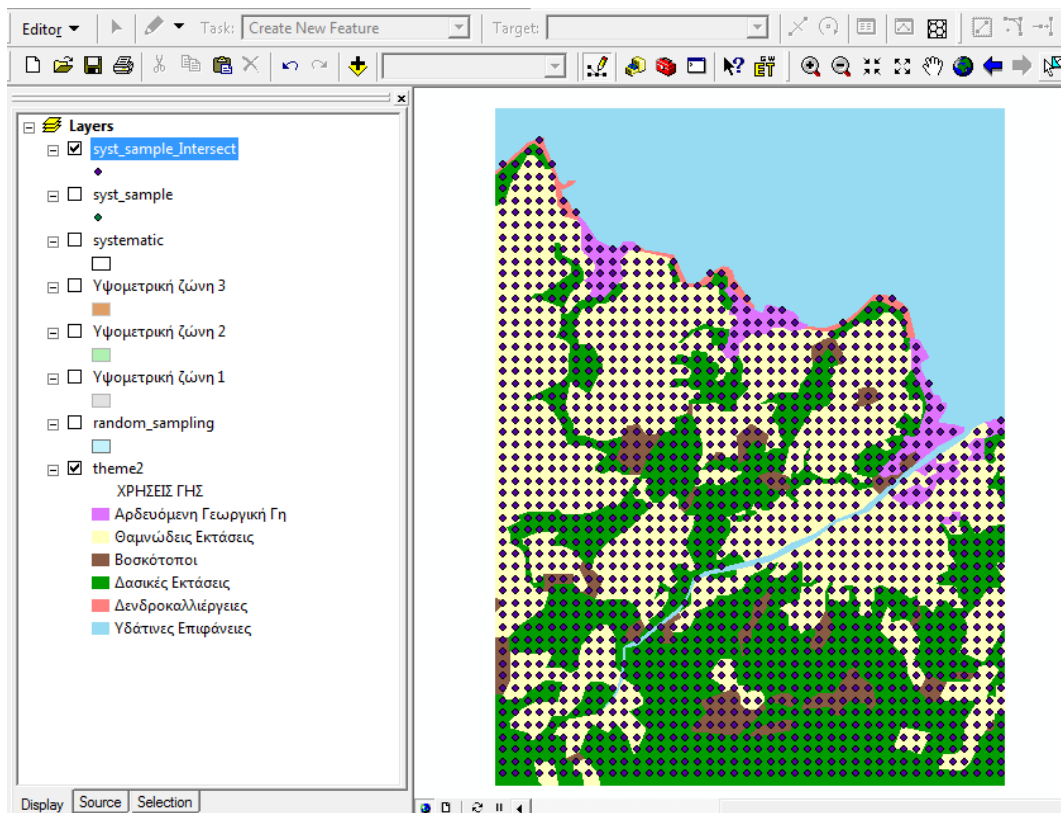
σχήμα 3.37: Δημιουργία κανονικώς κατανεμημένων σημείων στα όρια του επιπέδου systematic

- Τομή (intersect) του σημειακού επιπέδου που δημιουργήθηκε με το υπόβαθρο random_sampling. [arctoolbox > intersect]



σχήμα 3.38: Το εργαλείο intersect

Αποτέλεσμα της τομής είναι ένα νέο σημειακό επίπεδο που είναι πλέον προσδιορισμένο στο υπόβαθρο.



σχήμα 3.39: Συστηματική δειγματοληψία – απόσταση σημείων κατά X και Y: 50 m

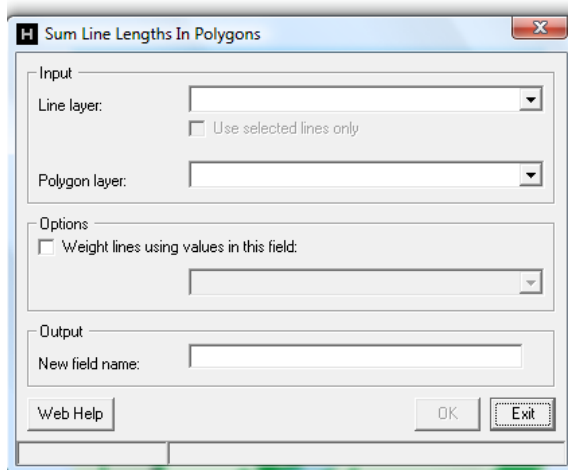
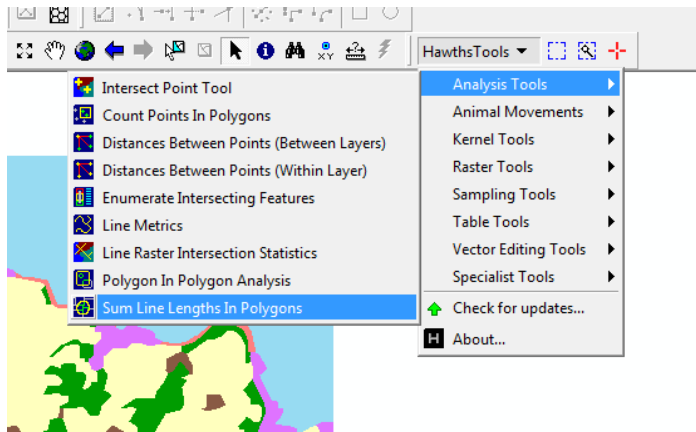
- Υπολογισμός σημείων ανά κατηγορία, με τον ίδιο τρόπο που έχει παρουσιαστεί και στην απλή τυχαία δειγματοληψία. (count points in polygon > summarize).

OID	LU_CODE	Count_LU_CODE	Sum_syst
0	100	7	57
1	200	19	818
2	300	26	83
3	400	20	703
4	500	1	0
5	700	3	15

σχήμα 3.40: σημεία ανά χρήση γης , 1676 σημεία

3.8 Εφαρμογή γραμμικής τυχαίας δειγματοληψίας

Για την εφαρμογή γραμμικών δειγματοληψιών η μπάρα Hawth tools έχει ένα χρήσιμο εργαλείο για τον υπολογισμό του συνολικού μήκους γραμμών που διέρχονται από κάθε πολύγωνο ενός θεματικού επιπέδου [Sum line lengths in polygons].

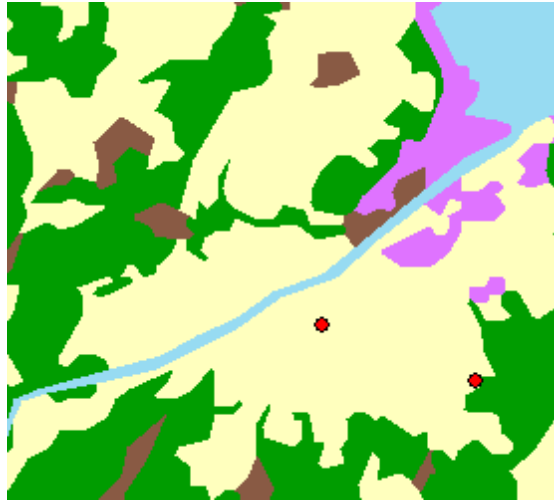


σχήμα 3.41: Παράθυρο διαλόγου Sum line lengths in polygons

Όπως έχει αναφερθεί και στο 1^ο κεφάλαιο στη περίπτωση της γραμμικής δειγματοληψίας, η αναζητούμενη πληροφορία καταγράφεται κατά μήκος των γραμμών.

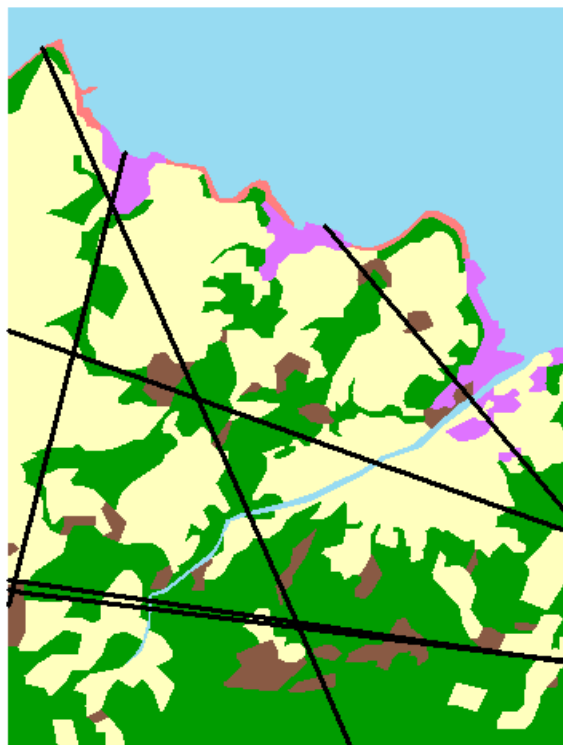
Για την εφαρμογή αυτής της μεθόδου ακολουθείται η παρακάτω διαδικασία:

- Η διαδικασία ξεκινά με την λήψη δύο τυχαίων σημείων του υποβάθρου που θα ορίσουν μία γραμμή δειγματοληψίας (σχήμα 3.42),



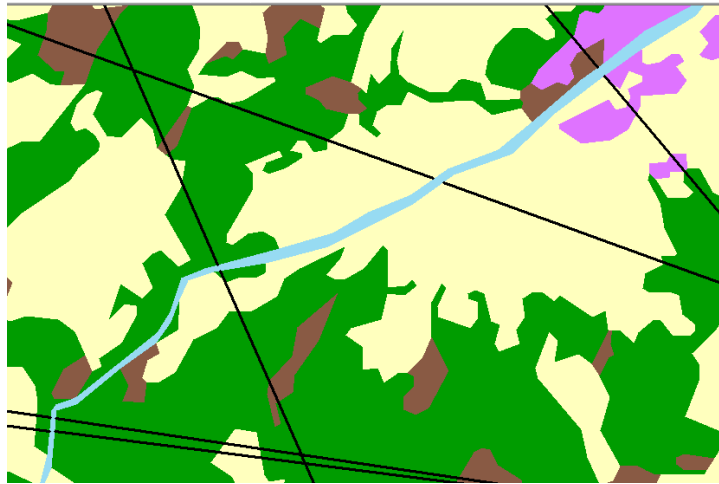
σχήμα 3.42: λήψη δύο τυχαίων σημείων

- Δημιουργείται ευθεία μεταξύ των δύο τυχαίων σημείων – με τον editor – και έπειτα προεκτείνεται η ευθεία στα όρια του δειγματοληπτικού υποβάθρου. Η δημιουργία της γραμμής γίνεται στο θεματικό επίπεδο transect που υπάρχει στον πίνακα περιεχομένων.
- Η παραπάνω διαδικασία αυτή επαναλαμβάνεται για το σύνολο των γραμμών που απαιτούνται.



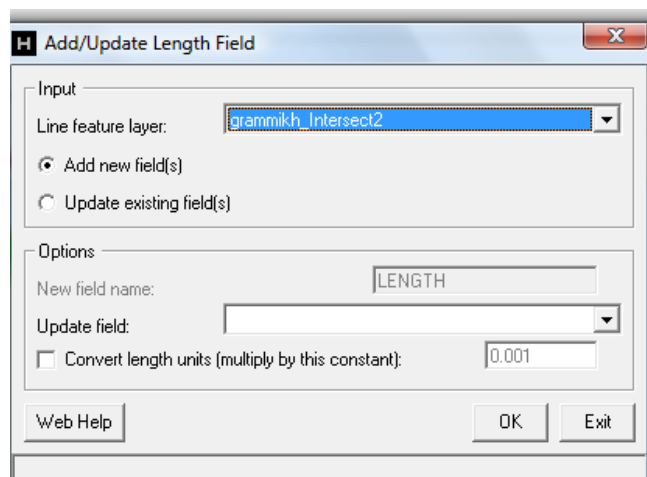
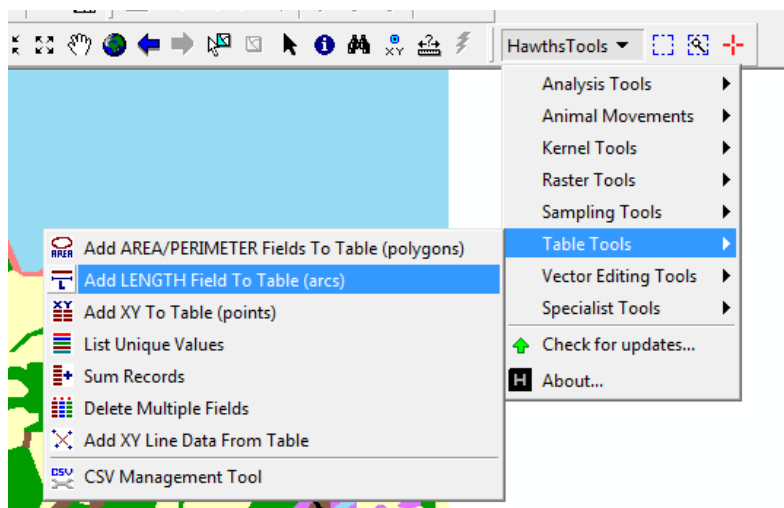
σχήμα 3.43: οι έξι γραμμές που παρήχθησαν

- Έπειτα γίνεται τομή (intersect - arctoolbox) του γραμμικού επιπέδου [transect] και του υποβάθρου (random sampling) με αποτέλεσμα τη δημιουργία ενός νέου γραμμικού σχηματικού αρχείου.



σχήμα 3.44: γραμμές δειματοληψίας μετά την τομή

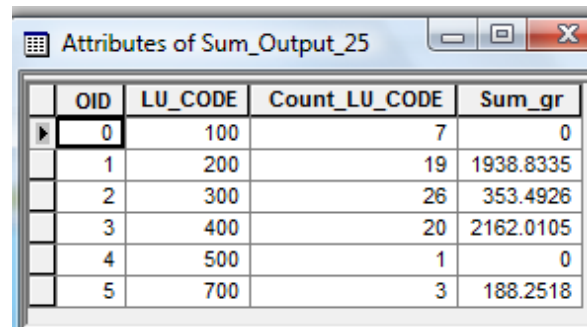
- Προσθήκη πεδίου στο πίνακα περιγραφικών χαρακτηριστικών των γραμμών στο οποίο αναγράφεται το μήκος της καθεμιάς. Από τη μπάρα Hawth's → Table Tools → Add length field to table.



σχήμα 3.45: Παράθυρο διαλόγου Add length field to table

- Η διαδικασία τελειώνει χρησιμοποιώντας το εργαλείο της μπάρας που παρουσιάστηκε στην αρχή της παραγράφου [sum line length in polygons] που ουσιαστικά είναι αντίστοιχο με το [count points in polygons] στη λογική βέβαια της γραμμικής δειγματοληψίας.

Για τον υπολογισμό τελικά του συνολικού μήκους γραμμών που περνά από κάθε χρήση γης γίνεται σύνοψη [summarize] στο πεδίο LU_CODE του πίνακα του επιπέδου theme_2 με το γνωστό από τη σημειακή δειγματοληψία τρόπο.



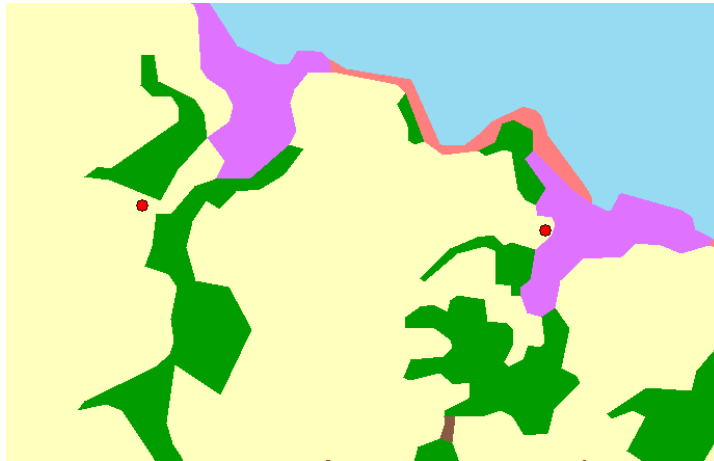
OID	LU_CODE	Count_LU_CODE	Sum_gr
0	100	7	0
1	200	19	1938.8335
2	300	26	353.4926
3	400	20	2162.0105
4	500	1	0
5	700	3	188.2518

σχήμα 3.46: Πίνακας εξόδου της δειγματοληψίας
Πεδίο [sum_gr]: το συνολικό μήκος γραμμών που διέρχονται από κάθε χρήση γης [LU_CODE]

Έχοντας υπολογίσει τα συνολικά μήκη των γραμμών (Sum_gr) ανά χρήση γης και γνωρίζοντας τα μήκη των γραμμών άρα και το άθροισμα των μηκών με μια απλή διαίρεση έχουμε τις εκτιμήσεις των ποσοστών μέσω της γραμμικής δειγματοληψίας.

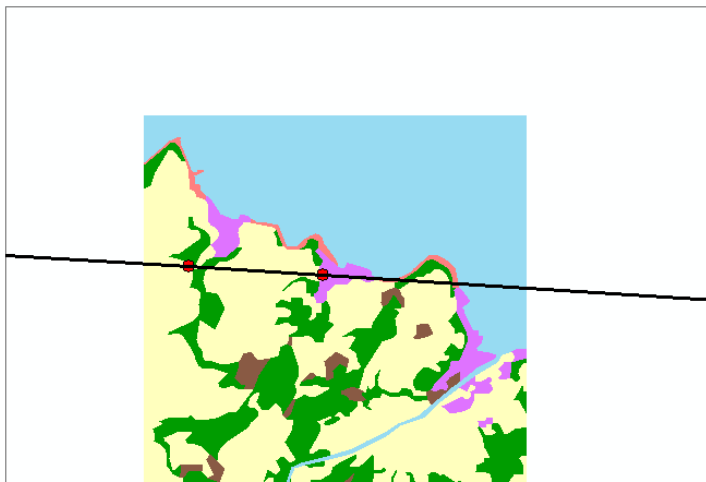
3.9 Συστηματική γραμμική δειγματοληψία

Η διαδικασία ξεκινά επιλέγοντας με τυχαίο τρόπο μία γραμμή, όπως και στη περίπτωση της τυχαίας γραμμικής. Για να γίνει κάτι τέτοιο λαμβάνονται δύο τυχαία σημεία στο υπόβαθρο της δειγματοληψίας (σχήμα 3.47).



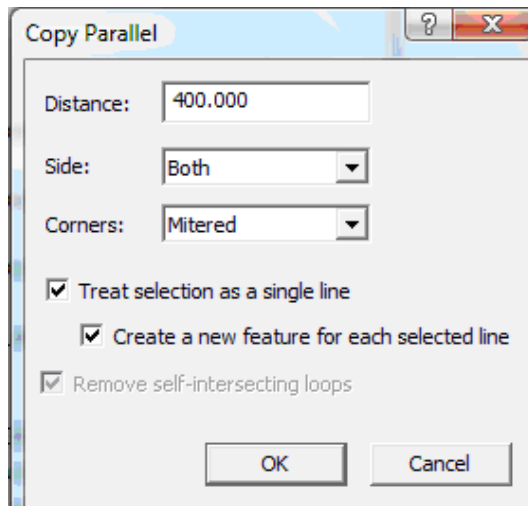
σχήμα 3.47: λήψη δύο τυχαίων σημείων

- Δημιουργούμε νέο σχηματικό αρχείο (γραμμικό) όπου θα δημιουργηθούν οι γραμμές δειγματοληψίας (π.χ. systematic transect)
- Χρησιμοποιώντας τον editor κατασκευάζεται η πρώτη γραμμή της δειγματοληψίας και προεκτείνεται (extend) στα όρια του επιπέδου extend (σχήμα 3.48). Για να γίνει η προέκταση θα πρέπει το επίπεδο (extend) να είναι επιλεγμένο.

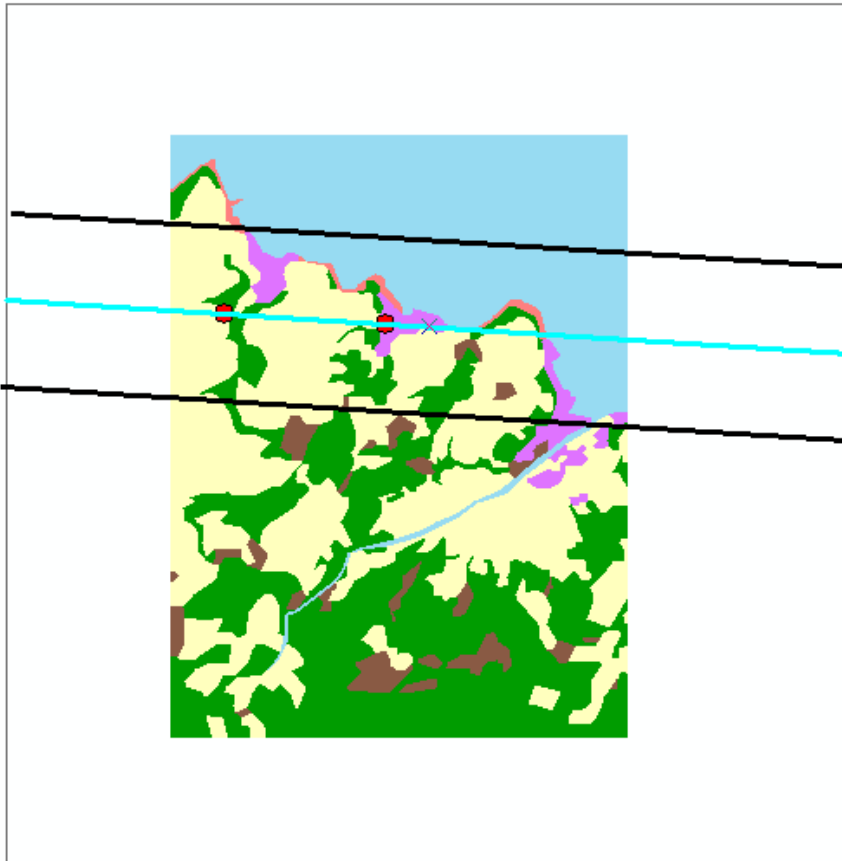


σχήμα 3.48: Δημιουργία της πρώτης γραμμής και προέκταση της

- Χρησιμοποιώντας και πάλι τον editor δημιουργούμε τις υπόλοιπες γραμμικές οντότητες που θα είναι παράλληλες της αρχικής, επιλέγοντας [editor > start editing > copy parallel]. Φυσικά θα πρέπει να είναι επιλεγμένη η παραπάνω γραμμική οντότητα που δημιουργήσαμε.



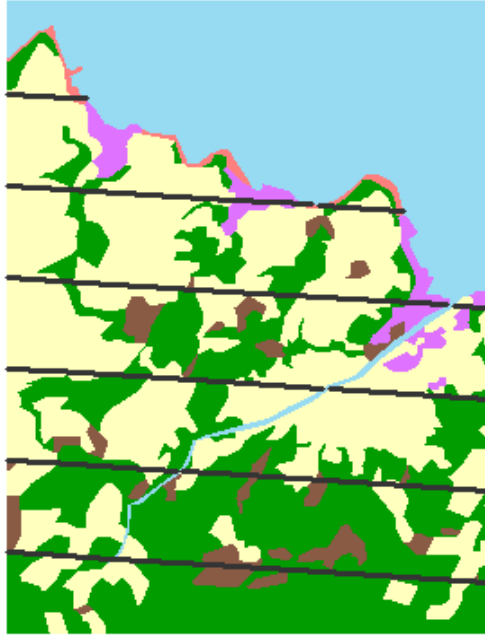
σχήμα 3.49: Δημιουργία παράλληλων γραμμικών οντοτήτων σε απόσταση 400μ



- Με αντίστοιχο τρόπο δημιουργούνται και οι υπόλοιπες γραμμές δειγματοληψίας
- Ακολουθεί η τομή του γραμμικού επιπέδου με το πλαίσιο της δειγματοληψίας
- Η συνέχεια της διαδικασίας είναι αντίστοιχη με αυτή της τυχαίας γραμμικής δηλαδή:

- Προσθήκη πεδίου στο πίνακα περιγραφικών χαρακτηριστικών των γραμμών στο οποίο αναγράφεται το μήκος της καθεμιάς. Από τη μπάρα Hawth's → Table Tools → Add length field to table.

- Άθροισμα των μηκών γραμμικών οντοτήτων που βρίσκονται σε κάθε πολύγωνο [Sum line lengths in polygons]



σχήμα 3.50: οι 6 γραμμές του συστηματικού γραμμικού δείγματος μετά την τομή των γραμμικών οντοτήτων με το υπόβαθρο

ΠΑΡΑΤΗΡΗΣΕΙΣ

Θέμα της διπλωματικής εργασίας ήταν η ανάπτυξη εκπαιδευτικού υλικού για την εφαρμογή δειγματοληπτικών μεθόδων σε περιβάλλον Γεωγραφικών Συστημάτων Πληροφοριών. Επιλέχθηκε να αναπτυχθούν οι πέντε βασικές μέθοδοι χωρικής δειγματοληψίας, η απλή τυχαία σημειακή, η στρωματοποιημένη σημειακή τυχαία, η συστηματική σημειακή δειγματοληψία, η γραμμική τυχαία και η γραμμική συστηματική δειγματοληψία. Η εφαρμογή τους έγινε σε έναν διανυσματικό χάρτη χρήσεων γης με σκοπό τον υπολογισμό των ποσοστών του εδάφους που αναλογούν σε κάθε χρήση.

Η εφαρμογή των δειγματοληψιών έγινε σε περιβάλλον γεωγραφικών συστημάτων πληροφοριών, έτσι ώστε να έχουμε στη διάθεση μας τα πραγματικά ποσοστά των χρήσεων των κατηγοριών του χάρτη – υπολογισμένα από τα εμβαδά - και να γίνεται έτσι η σύγκριση τους με αυτά που προκύπτουν από τις εφαρμογές των δειγματοληπτικών μεθόδων. Παράλληλα η εφαρμογή τους σε ΓΣΠ δίνει τη δυνατότητα στους σπουδαστές να χρησιμοποιήσουν βασικές λειτουργίες των συστημάτων αυτών, που πλέον αποτελούν αναπόσπαστο κομμάτι κάθε γεωγραφικής μελέτης.

Βασική παράμετρος κάθε δειγματοληψίας είναι ο υπολογισμός του απαιτούμενου μεγέθους του δείγματος, το οποίο στη περίπτωση εκτίμησης των αναλογιών κάποιου χαρακτηριστικού ενός πληθυσμού εξαρτάται από το επίπεδο εμπιστοσύνης που θα θέσουμε και προφανώς από το περιθώριο λάθους. Οι τιμές των δύο αυτών παραμέτρων καθορίζονται από τον μελετητή και σχετίζονται άμεσα με την ακρίβεια που θα παρουσιάζουν τα αποτελέσματα. Για το λόγο αυτό αποφασίστηκε οι δύο αυτοί παράμετροι να καθορίζονται στην αρχή της εφαρμογής κάθε δειγματοληψίας, είτε έπειτα από τη λήψη ενός δοκιμαστικού – προκαταρκτικού δείγματος, είτε χωρίς τη λήψη αυτού.

Το μέγεθος του δείγματος εξαιτίας των υψηλών απαιτήσεων ακρίβειας των αποτελεσμάτων (μικρό περιθώριο σφάλματος), για κάθε μέθοδο ήταν αρκετά μεγάλο. Την ανάγκη για λήψη μεγάλου δείγματος 'επέβαλε' και το γεγονός ότι υπάρχει κατηγορία χρήσης γης στο χάρτη – υπόβαθρο που αντιπροσωπεύει μόλις το 1% του συνόλου της έκτασης και ένα μικρό δείγμα υπήρχε μεγάλη πιθανότητα να μην αντιπροσωπεύει τη κατηγορία αυτή. Μια διαφορετική προσέγγιση θα ήταν η κατηγορία αυτή να συμπτυχτεί με κάποια άλλη.

Η φυσική στρωματοποίηση του υποβάθρου με βάση τις υψομετρικές ζώνες του χάρτη, δίνει αποτελέσματα αναφορικά με τα ποσοστά χρήσεων γης σε κάθε στρώμα, έτσι παράγεται μια επιπλέον πληροφορία και μπορεί να γίνει σύγκριση των ποσοστών των κατηγοριών χρήσεων σε κάθε υψομετρική ζώνη. Το γεγονός ότι δεν υπάρχει μεγάλη ομοιομορφία μέσα σε κάθε στρώμα και μεγάλη διαφοροποίηση μεταξύ των στρωμάτων δεν έχει ως αποτέλεσμα την επίτευξη μεγαλύτερης ακρίβειας σε σχέση με τις άλλες μεθόδους με μικρότερο ή το ίδιο μέγεθος δείγματος.

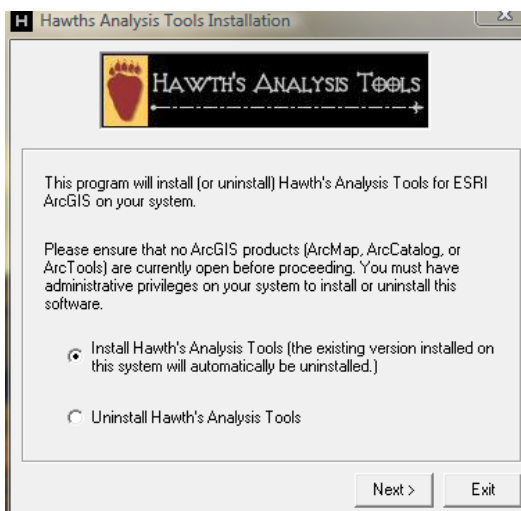
ΠΑΡΑΡΤΗΜΑ

Εγκατάσταση Hawth's Analysis Tools

Για την πραγματοποίηση των δειγματοληψιών γίνεται εγκατάσταση μιας πρόσθετης επέκτασης (extension) που ονομάζεται Hawth's Tools. Η επέκταση αυτή προσφέρει μια σειρά εργαλείων χωρικής ανάλυσης και παρέχει τη δυνατότητα σε μη εξοικειωμένους χρήστες να πραγματοποιούν σύνθετες λειτουργίες απλούστερα.

Αρχικά γίνεται λήψη του συμπιεσμένου αρχείου από την ιστοσελίδα που διατίθεται δωρεάν [www.spatial ecology.com/htools/download.php].

Για την εγκατάσταση αφού αποσυμπεστεί το αρχείο επιλέγεται το [htools_setup.exe]



Η εγκατάσταση συνεχίζεται με τη βοήθεια του παραθύρου που εμφανίζεται.

Μετά την επιτυχή εγκατάσταση της επέκτασης, η ενεργοποίηση της εργαλειοθήκης γίνεται από το μενού επιλογών του arcmap και συγκεκριμένα από το μενού [**tools**].

Επιλέγοντας [**customize**], στην καρτέλα [**toolbars**] μαρκάρεται η επιλογή [**Hawth's Tools**] και έπειτα [**close**]. Εάν δεν έχει εμφανιστεί ήδη η νέα εργαλειοθήκη, κάνοντας δεξί κλικ στο γκρι πλαίσιο – όπου τοποθετούνται οι μπάρες εργαλείων, μαρκάρεται η επιλογή Hawth's Tools. Πλέον, δίπλα σε κάποια από τις ήδη υπάρχουσες εργαλειοθήκες έχει τοποθετηθεί και η νέα μπάρα, έτοιμη για χρήση.

Το λογισμικό πακέτο ArcGIS 9.X

Για την ανάπτυξη της άσκησης χρησιμοποιήθηκε το λογισμικό πακέτο ArcGIS 9.2 της ESRI. Το πακέτο αυτό είναι ένα σύγχρονο λογισμικό στο χώρο των Γεωγραφικών Συστημάτων Πληροφοριών, αποτελεί ένα κλιμακωτό λογισμικό σύστημα για δημιουργία, διαχείριση, ενοποίηση, ανάλυση και διασπορά γεωγραφικών δεδομένων για κάθε οργανισμό – από ένα άτομο έως ένα παγκόσμιο καταναμημένο δίκτυο ανθρώπων. Η έκδοση αυτή αποτελεί ένα κατανοητό και ολοκληρωμένο σύστημα, σχεδιασμένο για να ικανοποιεί τις ανάγκες των περισσότερων χρηστών γεωγραφικών συστημάτων πληροφοριών.

Το λογισμικό παρουσιάζει κάποια βασικά χαρακτηριστικά:

- Είναι εύκολο στη χρήση – προσφέρει δυνατότητες γενικής χαρτογραφίας, ανάλυσης και διαχείρισης δεδομένων.
- Είναι ιδιαίτερα λειτουργικό – ενσωματώνει δυνατότητες επέμβασης στα γεωγραφικά δεδομένα, βελτίωση στη διαχείριση των δεδομένων και σύνθετη χωρική ανάλυση.
- Είναι κλιμακωτό – όλα τα επιμέρους επίπεδα του λογισμικού χρησιμοποιούν τον ίδιο πυρήνα εφαρμογών, τις ίδιες λειτουργικές αρχές και το ίδιο user interface.
- Έχει δυνατότητες Web – δίνει τη δυνατότητα στους χρήστες να χρησιμοποιούν το διαδίκτυο για να αποκτούν άμεση πρόσβαση σε δεδομένα και υπηρεσίες.
- Είναι φιλικό προς το χρήστη – ο χρήστης έχει τη δυνατότητα παραγωγής δικών του επεκτάσεων, εφαρμογών, εργαλείων (scripts), σε διάφορες γλώσσες προγραμματισμού (πχ Visual Basic) ώστε να πετυχαίνει τη βέλτιστη απόδοση στην εργασία του.

Ο πυρήνας του λογισμικού είναι τα προγράμματα ArcMap, ArcCatalog και ArcToolbox.

Το ArcMap χρησιμοποιείται για τη πλήρη διαχείριση των χωρικών δεδομένων και τη δημιουργία χαρτών. Πιο συγκεκριμένα, δίνει τη δυνατότητα να δημιουργηθούν χάρτες από επίπεδα χωρικής πληροφορίας, να αναλυθούν χωρικές σχέσεις και να επιλεγούν μέσα από αναζητήσεις χωρικά και μη χωρικά στοιχεία. Επίσης μπορούν να σχεδιαστούν και να δημιουργηθούν διαφορετικές απεικονίσεις ενός χάρτη, αλλάζοντας χρώματα και συμβολισμούς.

Το ArcCatalog χρησιμοποιείται για τον εντοπισμό, την οργάνωση και τη διαχείριση αρχείων χωρικών δεδομένων. Ουσιαστικά εκτελεί παρόμοιες λειτουργίες με έναν τυπικό explorer, προσανατολισμένο βέβαια στα γεωγραφικά δεδομένα.

Το ArcToolbox διαθέτει εργαλεία μετατροπής αρχείων γεωγραφικών δεδομένων και για Geoprocessing. Geoprocessing (Γεωεπεξεργασία) είναι ο μηχανισμός μέσω του οποίου επιτυγχάνεται η μετάβαση από στοιχεία σε πληροφορία (θεμελιώδης έννοια σε οποιαδήποτε γεωγραφική μελέτη) με τη χρήση του ArcGIS. Αποτελεί μια σειρά από εργαλεία για τη διαχείριση και ανάλυση των χωρικών δεδομένων. Τα δύο βασικά τμήματα του Geoprocessing είναι τα εργαλεία του ArcToolbox και οι αναλυτικές διαδικασίες του ModelBuilder.

Το ArcToolbox στη πλήρη έκδοσή του (έκδοση ArcInfo) περιλαμβάνει περίπου 200 εργαλεία για γεωεπεξεργασία, μετατροπή δεδομένων, διαχείριση χαρτογραφικών φύλλων, αλληλεπίθεση επιπέδων, χαρτογραφικών προβολών κ.α.

Οι λειτουργίες και οι δυνατότητες του κάθε προγράμματος αυξάνονται καθώς ανεβαίνουμε από το χαμηλότερο στο υψηλότερο επίπεδο του λογισμικού πακέτου, από το ArcView, στο ArcEditor και τέλος στο ArcInfo.

Το ArcMap αποτελεί τη κεντρική εφαρμογή του λογισμικού πακέτου ArcGIS desktop 9.X. Χρησιμοποιείται για όλε στις διεργασίες που βασίζονται σε χάρτες, συμπεριλαμβανομένης και της επέμβασης στα χαρτογραφικά στοιχεία (editing). Οι χάρτες έχουν μια εικόνα εξόδου (page layout) που εμφανίζει το χάρτη με τα επιθυμητά θεματικά επίπεδα, το υπόμνημα, τη κλίμακα, τον προσανατολισμό και άλλα δευτερεύοντα γραφικά στοιχεία. Το ArcMap προσφέρει και τη γεωγραφική όψη του χάρτη (geographic data view), που εκεί μπορεί να εκτελεστεί μια πληθώρα εφαρμογών των γεωγραφικών συστημάτων πληροφοριών.

Ο χρήστης έχει δυνατότητες όπως:

- να δημιουργήσει και να αλληλεπιδράσει με χάρτες,
- να συνθέσει και να αναλύσει γεωγραφικά δεδομένα,
- να θέσει ερωτήματα στα χωρικά δεδομένα του χάρτη, να βρει και να αποκαλύψει τις μεταξύ τους χωρικές σχέσεις,
- να συμβολίσει τις γεωγραφικές πληροφορίες με μια μεγάλη ποικιλία τρόπων και συμβολισμών,
- να δημιουργήσει γραφήματα και εκθέσεις – αναφορές ώστε να γίνεται δυνατή η επικοινωνία με άλλους χρήστες και για να γίνονται πιο σαφείς και διακριτές οι διάφορες απεικονίσεις και δεδομένα.
- να δημιουργήσει τελικές εικόνες εξόδου – χάρτες και στην οθόνη του Η/Υ να παρουσιάζεται αυτό ακριβώς που θα εκτυπωθεί.

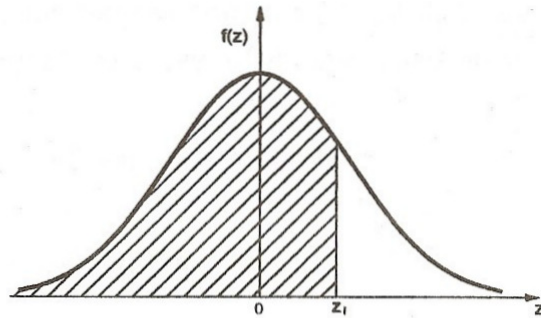
Το ArcCatalog όπως προαναφέρθηκε είναι εφαρμογή διαχείρισης των γεωγραφικών δεδομένων. Μοιάζει σε εμφάνιση και λογική με τον Explorer των Windows.

Μέσω του ArcCatalog ο χρήστης έχει τη δυνατότητα:

- να διαχειριστεί αρχεία γεωγραφικών δεδομένων (αντιγραφή, μετακίνηση, μετονομασία, διαγραφή) και φακέλους,
- να κάνει επισκόπηση (preview) των αρχείων γεωγραφικών δεδομένων,
- να κάνει επισκόπηση ή τροποποίηση (edit) των μεταδεδομένων (=δεδομένα για τα δεδομένα) που συνοδεύουν τα αρχεία γεωγραφικών δεδομένων ή και να τα δημιουργήσει εξ αρχής,
- να εισάγει απευθείας δεδομένα στο ArcMap με drag & drop,
- να κάνει αναζήτηση δεδομένων (search) τόσο στον δίσκο του υπολογιστή ή ακόμα και στο διαδίκτυο,
- να συνδεθεί με προηγμένες υπηρεσίες παροχής γεωγραφικών δεδομένων μέσω διαδικτύου.

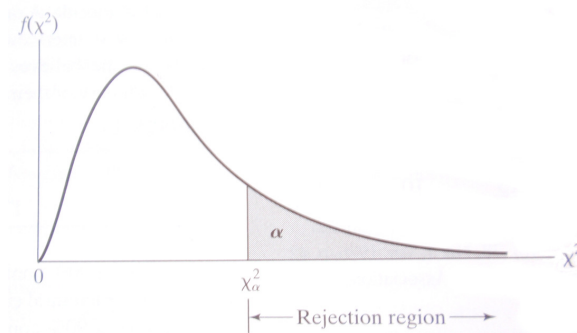
Πίνακας Κανονικής Κατανομής

$$P(-\infty < z < z_i) = \int_{-\infty}^{z_i} f(z) dz = F(z_i)$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

Πίνακας Κρίσιμων τιμών του χ^2



Βαθμοί ελευθερίας	$\chi^2_{0.995}$	$\chi^2_{0.950}$	$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$
1	0.0000393	0.0039321	2.70554	3.84146	5.02389
2	0.0100251	0.102587	4.60517	5.99147	7.37776
3	0.0717212	0.351846	6.25139	7.81473	9.34840
4	0.206990	0.710721	7.77944	9.48773	11.1433
5	0.411740	1.145476	9.23635	11.0705	12.8325

ΠΙΝΑΚΕΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΔΕΙΓΜΑΤΟΛΗΨΙΩΝ

Παρακάτω παρατίθενται πίνακες αποτελεσμάτων εκτιμήσεων των ποσοστών χρήσεων γης του χάρτη λαμβάνοντας διαφορετικά μεγέθη δείγματος και χρησιμοποιώντας τις διάφορες μεθόδους επιλογής τους. Για κάθε δείγμα δίνονται τα αποτελέσματα τριών επαναλήψεων της διαδικασίας.

<u>Πραγματικά Ποσοστά Χρήσεων Γης</u>	
Αρδευόμενη γεωργική γη	3.6 %
Θαμνώδεις εκτάσεις	48.9 %
Βοσκότοποι	5 %
Δασικές εκτάσεις	41.6 %
Δενδροκαλλιέργειες	0.9 %

Σημειακή τυχαία δειγματοληψία

Μέγεθος δείγματος : 30 σημεία

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλλιέργειες
1	3.3 %	56.7 %	6.7 %	33.3 %	0
2	10 %	50 %	10 %	30 %	0
3	3.3 %	43.3 %	0 %	53.4 %	0

Μέγεθος δείγματος : 600 σημεία

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλλιέργειες
1	3.2 %	48.2 %	4.7 %	43.2 %	0.8 %
2	4.2 %	47.2 %	3.5 %	43.6 %	1.5 %
3	4.3 %	51.5 %	4 %	40 %	0.5 %

Μέγεθος δείγματος 2401 σημεία

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλλιέργειες
1	3.2 %	47.7 %	5.8 %	42.2 %	1.1 %
2	3.7 %	48.2 %	4.7 %	42.2 %	1.2 %
3	3.3 %	49.7 %	4.9 %	41.5 %	0.6 %

Στρωματοποιημένη τυχαία σημειακή δειγματοληψία

Πραγματικά Ποσοστά Χρήσεων γης ανά Στρώμα

	1 ^ο στρώμα	2 ^ο στρώμα	3 ^ο στρώμα
Αρδευόμενη γεωργική γη	13.6 %	0.2 %	0 %
Θαμνώδεις εκτάσεις	54.5 %	69.9 %	19.6%
Βοσκότοποι	2.5 %	4.6 %	7.3%
Δασικές εκτάσεις	25.8 %	25.3%	73.1%
Δενδροκαλλιέργειες	3.6 %	0 %	0 %

Μέγεθος δείγματος : 30 σημεία

1^ο στρώμα (8)

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλλιέργειες
1	87.5 %	12.5 %	0 %	0 %	0 %
2	50 %	37.5 %	0 %	12.5 %	0 %
3	25 %	62.5 %	12.5 %	0 %	0 %

2^ο στρώμα (12)

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλλιέργειες
1	0 %	66.7 %	8.3 %	25 %	0 %
2	0 %	58.3 %	0 %	41.7 %	0 %
3	0 %	58.3 %	0 %	41.7 %	0 %

3^ο στρώμα (10)

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλλιέργειες
1	0 %	30 %	10 %	60 %	0 %
2	0 %	0 %	10 %	90 %	0 %
3	0 %	20 %	0 %	80 %	0 %

Μέγεθος δείγματος 2401 σημεία

1^ο στρώμα (619)

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλλιέργειες
1	14.2 %	55.1%	3.4 %	23.6 %	3.7 %
2	13.2%	54.8%	2.1%	26.5%	3.4%
3	11.3%	54.4%	2.3%	27.5%	4.5%

2^ο στρώμα (969)

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλλιέργειες
1	0.1%	71.6%	5.0%	23.3%	0%
2	0%	70.2%	5.0%	24.8%	0%
3	0%	71%	5.8%	23.2%	0%

3^ο στρώμα (813)

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλλιέργειες
1	0%	18.6%	6.5%	74.9%	0%
2	0%	19.2%	7.3%	73.5%	0%
3	0%	17.8%	7.6	74.5%	0%

Συστηματική τυχαία σημειακή δειγματοληψία

Απόσταση μεταξύ των σημείων 350μ

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλ- λιέργειες
1 (33 σημεία)	3.0%	39.5%	0%	54.5%	3.0%
2 (35 σημεία)	0%	42.9%	2.9%	51.3%	2.9%
3 (31 σημεία)	3.2%	38.7%	9.7%	48.4%	0%

Απόσταση μεταξύ των σημείων 150μ

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλ- λιέργειες
1 (185 σημεία)	5.4%	48.7%	4.3%	40.0%	1.6%
2 (193 σημεία)	4.7%	51.3%	5.2%	38.3%	0.5%
3 (194 σημεία)	3.1%	47.9%	4.2%	44.8%	0%

Απόσταση μεταξύ των σημείων 40μ

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλ- λιέργειες
1 (2633 σημεία)	3.2%	49.0%	5.1%	41.7%	1%

Γραμμική τυχαία δειγματοληψία

6 γραμμές

Επανάληψη Δειγματοληψίας	Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλ- λιέργειες
1	3.6%	39.8%	7.8%	48.7	0.1%
2	3.1%	42.4%	5.9%	48.4%	0.2%
3	2.6%	40.6%	7.3%	49.3%	0.2%

Γραμμική τυχαία συστηματική δειγματοληψία (6 γραμμές)

Αρδευόμενη γεωργική γη	Θαμνώδεις εκτάσεις	Βοσκότοποι	Δασικές εκτάσεις	Δενδροκαλ- λιέργειες
4.8%	53.7%	3.7%	36%	1.8%

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Αγατζα – Μπαλοδήμου Α.Μ, 2004, *Θεωρία Σφαλμάτων και Συνορθώσεις I (σημειώσεις από τις παραδόσεις του μαθήματος)*, Ε.Μ.Π.
- Κάβουρας Μ., 2003, *Αρχές Γεωπληροφορικής και Συστήματα Γεωγραφικών Πληροφοριών*, Ε.Μ.Π.
- Κοκολάκης Γ, Σπηλιώτης Ι, 1999, *Εισαγωγή στη Θεωρία Πιθανοτήτων και Στατιστική*, Συμεών
- Κουτσόπουλος Κ., 2005, *Γεωγραφικά Συστήματα Πληροφοριών*, Παπασωτηρίου
- Κουτσόπουλος Κ., Ανδρουλακάκης Ν., 2005, *Εφαρμογές του λογισμικού ArcGIS 9x με απλά λόγια*, Παπασωτηρίου
- Κουτσόπουλος Κ., 2006, *Ανάλυση Χώρου: Θεωρία, Μεθοδολογία και Τεχνικές*, Διηνεκές
- Μπένος Β, 1991, *Μέθοδοι και Τεχνικές Δειγματοληψίας*, Εκδόσεις Σταμούλης Πειραιάς
- Τζιαφέτας Γ, 1991, *Θεωρία και Τεχνική Δειγματοληψίας*, τομέας εκδόσεων Ελληνικού κέντρου Παραγωγικότητας
- Bonham G., 1994, *Geographic Information Systems for Geoscientists (modelling with GIS)*, Pergamon
- Cochran W., 1977, *Sampling Techniques (3rd edition)*, John Wiley & sons
- Dixon C., Leach B., 1977 *Sampling Methods for Geographical Research*, Catmog
- Mendenhall W., Sincich T., 2007, *Statistics for Engineering and the Sciences (fifth edition)*, Prentice Hall
- Monroe Ch, 1993, *An introduction to statistical problem solving in geography*, Wm.C.Brown Publishers, 1993
- Shaw G. and Wheeler D., 1985, *Statistical Techniques in Geographical Analysis*, John Wiley & sons

Κυριότερες πηγές στο διαδίκτυο

- <http://support.esri.com/>
- <http://www.spatial ecology.com/>