

Χωρική Δειγματοληψία

5.1. ΕΙΣΑΓΩΓΗ

Η δειγματοληψία είναι ένα βασικό ερευνητικό εργαλείο της γεωγραφικής επιστήμης (όπως και του συνόλου σχεδόν των επιστημών). Μολονότι τα δείγματα κατά την πολύχρονη ιστορία της γεωγραφικής έρευνας έχουν χρησιμοποιηθεί, είτε για να αποκτηθεί μια γενική εντύπωση κάποιας περιοχής είτε για να γίνουν εκτιμήσεις για ορισμένα από τα χαρακτηριστικά της, εντούτοις η προσοχή και οι προσπάθειες της μοντέρνας γεωγραφίας έχουν στραφεί στη δεύτερη προσέγγιση, δηλαδή στη στατιστική χρήση των δειγμάτων. Έτσι, αυτό το κεφάλαιο αναφέρεται στις βασικές έννοιες, στους τρόπους και στα προβλήματα της **στατιστικής χωρικής δειγματοληψίας**.

Κάθε συζήτηση γύρω από τη στατιστική δειγματοληψία οφείλει να ξεκαθαρίζει τις έννοιες του **πληθυσμού** και του **δείγματος**. Στη στατιστική, η λέξη πληθυσμός έχει μια τεχνική έννοια που δεν έχει καμιά σχέση με τον ανθρώπινο πληθυσμό. Ας υποθεθεί ότι υπάρχει ανάγκη να εξακριβωθούν διάφορα χαρακτηριστικά μιας περιοχής (π.χ. το μέγεθος των αγροτεμαχίων της και το είδος της καλλιέργειάς τους). Αν ήταν δυνατή η καταγραφή καθενός από αυτά τα χαρακτηριστικά, θα υπήρχαν δύο πληθυσμοί: ο ένας θα αποτελούσε το σύνολο των **μετρήσεων** (*measurements*) για το μέγεθος όλων των αγροτεμαχίων της περιοχής και ένας δεύτερος το σύνολο με **απαριθμήσεις** (*counts*) για τα είδη των καλλιεργειών στην περιοχή. Στη στατιστική, επομένως, πληθυσμός είναι ένα σύνολο από μετρήσεις ή απαριθμήσεις που προέρχονται από όλα τα αντικείμενα που έχουν ένα ή περισσότερα κοινά χαρακτηριστικά. Στο προηγούμενο παράδειγμα το κοινό χαρακτηριστικό αναφερόταν στο αντικείμενο «**αγροτεμάχια μιας ορισμένης περιοχής**». Στην ειδική περίπτωση της γεωγραφίας, **γεωγραφικός πληθυσμός** είναι ένα σύνολο μετρήσεων ή απαριθμήσεων για αντικείμενα με κάποια κοινή γεωγραφική-χωρική ιδιότητα.

Αναμφισβήτητα, ο καλύτερος τρόπος μελέτης των ιδιοτήτων ενός γεωγραφικού πληθυσμού είναι η μελέτη κάθε αντικειμένου χωριστά. Συχνά όμως, δεν είναι

αναγκαίο αλλά ούτε και πρακτικό να αναζητούνται μετρήσεις ή απαριθμήσεις για ολόκληρο τον πληθυσμό. Δηλαδή, αν αναζητείται το μέσο μέγεθος των αγροτεμαχίων ενός νομού και υπάρχουν αρκετές χιλιάδες αγροτεμάχια, υπάρχουν διαδικασίες για να αποκτηθεί μια πολύ καλή, και για τους περισσότερους σκοπούς, απόλυτα ικανοποιητική προσέγγιση: μετρώντας το μέγεθος ενός μικρού αριθμού απ' όλα τα αγροτεμάχια και εξοικονομώντας με τον τρόπο αυτό χρόνο, χρήματα και προσπάθεια. Τα στοιχεία που συλλέγονται μ' ένα τέτοιο τρόπο, αποτελούν ένα **δείγμα**, που μπορεί να οριστεί απλά ως ένα τμήμα του πληθυσμού.

Πρέπει να σημειωθεί, όμως, ότι στη στατιστική ένα δείγμα, όπως και ένας πληθυσμός, αποτελεί συλλογή στοιχείων, όχι αντικειμένων. Δηλαδή, όταν αναφέρεται ότι «ο πληθυσμός αποτελείται από όλα τα νοικοκυριά της Αθήνας», βασικά αυτό που εννοείται είναι ότι «ο πληθυσμός αποτελείται από ορισμένα στοιχεία που αναφέρονται σε όλα τα νοικοκυριά της Αθήνας».

Ένα δείγμα, επομένως, αντικαθιστά τον πληθυσμό για να εξοικονομηθεί χρόνος, χρήματα και προσπάθεια ή όταν μετρήσεις σε ολόκληρο τον πληθυσμό δεν είναι δυνατό ή πρακτικό να γίνουν. Για να είναι, όμως, ένα δείγμα αξιόπιστος αντικαταστάτης του πληθυσμού, πρέπει να είναι **αντιπροσωπευτικό**, δηλαδή πρέπει να εκφράζει τις διαφοροποιήσεις του πληθυσμού όσο γίνεται πιο πιστά. Επομένως, αναφορικά με το προηγούμενο παράδειγμα, η αναλογία των αγροτεμαχίων που έχουν μικρό, μεσαίο και μεγάλο μέγεθος πρέπει να είναι περίπου η ίδια όπως και στον πληθυσμό. Για να επιτευχθεί αυτό, είναι σημαντικό να αποφεύγεται η **μεροληπτικότητα** (*bias*) που είναι αποτέλεσμα της επιλογής για δειγματοληψία μη αντιπροσωπευτικών τμημάτων του πληθυσμού.

Αν τα δείγματα επιλεγούν με βάση την προσωπική κρίση του ερευνητή, είναι σχεδόν σίγουρο ότι θα είναι μεροληπτικά. Υποκειμενική δειγματοληψία είναι αδύνατο να γίνει χωρίς υποσυνείδητες επιλογές. Είναι έτσι αναπόφευκτο από έναν ερευνητή να μην παρατηρεί φυτά με ωραία λουλούδια περισσότερο από φυτά χωρίς λουλούδια και να μην ερωτά αναλογικά περισσότερους ωραίους αντιπροσώπους του άλλου φύλου κ.λπ. Αλλά και αν ακόμη υπάρχει υποκειμενική επιλογή χωρίς μεροληπτικότητα, αυτή δυστυχώς δεν μπορεί ν' αποδειχτεί. Επομένως, είναι αναγκαίο να ακολουθούνται αντικειμενικές διαδικασίες που θα μειώνουν στο ελάχιστο το υποκειμενικό στοιχείο αυτών των επιλογών.

5.2. ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΟ ΥΠΟΒΑΘΡΟ

Εκεί που κυρίως απαιτείται κρίση είναι στην επιλογή του κατάλληλου δειγματοληπτικού υποβάθρου, που είναι ο στατιστικός πληθυσμός από όπου λαμβάνεται το δείγμα. Τα αποτελέσματα μιας δειγματοληψίας έχουν εφαρμογή, αναφέρονται δηλαδή αποκλειστικά και μόνο στον πληθυσμό από τον οποίο έχει ληφθεί το δείγμα, γιατί πολύ συχνά κάτω από τις συνθήκες που επικρατούν, ο πραγματικός πληθυσμός δεν προσφέρεται για δειγματοληψία. Για παράδειγμα, στην Ελλάδα το λεγόμενο 5% δείγμα της στατιστικής υπηρεσίας, ενώ αναφέρεται στο 5% των κατοίκων της χώρας μας, ουσιαστικά είναι ένα δείγμα 5% των νοικοκυριών της, γιατί είναι σχεδόν αδύνατο να πραγματοποιηθεί μια στατιστικά άμεμπτη δειγματοληψία με βάση τα άτομα.

Η επιλογή ενός δειγματοληπτικού υποβάθρου το οποίο να περιέχει ολόκληρο το φάσμα των μετρήσεων και όλα τα είδη των απαριθμήσεων του υπό εξέταση πληθυσμού, είναι το πρώτο και βασικό βήμα σε μια δειγματοληψία. Εάν το υπόβαθρο δεν ανταποκρίνεται πλήρως στον παραπάνω περιορισμό, ο ερευνητής οφείλει να καθορίσει, για τη συγκεκριμένη μελέτη του, τις συνθήκες κάτω από τις οποίες μπορεί να αποδεχθεί τυχόν απόκλιση ή αν θα πρέπει να διαφοροποιήσει το υπάρχον ή να επιλέξει ένα άλλο δειγματοληπτικό υπόβαθρο. Κι αυτό γιατί το δειγματοληπτικό υπόβαθρο μπορεί να αποτελέσει ένα σημαντικό παράδειγμα μεροληπτικότητας του δείγματος, ανάλογα με το πόσο αντιπροσωπευτικό δείγμα είναι του πραγματικού πληθυσμού που μελετάται. Για παράδειγμα, ο κατάλογος όσων συνδέθηκαν στο διαδίκτυο τα δύο τελευταία χρόνια μπορεί να αποτελέσει ένα καλό δειγματοληπτικό υπόβαθρο, αν ο πραγματικός πληθυσμός είναι οι οικογένειες που διαθέτουν Η/Υ· θα ήταν μεροληπτικό υπόβαθρο αν χρησιμοποιείτο για να αντιπροσωπεύσει τα νοικοκυριά της Ελλάδας. Με τον ίδιο τρόπο, ακτινωτές οδεύσεις από το κέντρο της πόλης μπορούν να αποτελέσουν ένα χρήσιμο δειγματοληπτικό υπόβαθρο για την εκτίμηση της σχέσης μεταξύ απόστασης από το κέντρο και κάποιας άλλης μεταβλητής. Θα ήταν μεροληπτικό, όμως, σε όφελος των περιοχών κοντά στο κέντρο, αν χρησιμοποιείτο για την εκτίμηση της αναλογίας των διαφόρων χρήσεων γης, γιατί ο χώρος μεταξύ των ακτινωτών οδεύσεων αυξάνει όσο αυξάνει η απόσταση από το κέντρο.

Πρέπει να σημειωθεί, όμως, ότι η επιλογή του δειγματοληπτικού υποβάθρου είναι σχεδόν πάντοτε μια διαδικασία επιλογής μεταξύ μη βέλτιστων εναλλακτικών. Για παράδειγμα, αν τα υπάρχοντα διαθέσιμα (π.χ. χρόνος, χρήματα κ.λπ) είναι περιορισμένα, η χρησιμοποίηση ενός ήδη υπάρχοντος υποβάθρου, είναι προτιμότερη

από τη διαφοροποίηση ή την αλλαγή του, με ταυτόχρονη διοχέτευση των διαθέσιμων σε άλλες ενέργειες της συλλογής στοιχείων. Κι αυτό γιατί η δημιουργία ενός δειγματοληπτικού υποβάθρου είναι σύνθετη, χρονοβόρα και δαπανηρή διαδικασία, ακόμη και στην περίπτωση δημιουργίας ενός υποβάθρου με τη μορφή απλού καταλόγου.

Μολονότι οι γεωγράφοι ενδιαφέρονται για χωρικά δείγματα και, επομένως, έχουν ιδιαίτερη ανάγκη καθορισμού χωρικών δειγματοληπτικών υποβάθρων, εντούτοις συχνά κάνουν χρήση και μη χωρικών δειγματοληπτικών υποβάθρων. Γενικά, η φύση του δειγματοληπτικού υποβάθρου εξαρτάται από το αντικείμενο της έρευνας. Αν το ενδιαφέρον εστιάζεται σε ένα χαρακτηριστικό για μια **περιοχή** ως **σύνολο** (π.χ. ο αριθμός των φαρμακείων στο Βόλο, με την προϋπόθεση ότι η περιοχή του Βόλου είναι σαφώς ορισμένη), τότε η χρήση ενός μη χωρικού υποβάθρου είναι ικανοποιητική (π.χ. κατάλογος από το εμπορικό επιμελητήριο). Αν όμως η μελέτη αφορά τη χωρική διάσταση μιας περιοχής στο σύνολό της (την έκταση που καταλαμβάνουν τα φαρμακεία του Βόλου) ή για χωρικές διαφοροποιήσεις **μέσα σε μια περιοχή** (π.χ. διαφορές στην έκταση που καταλαμβάνουν τα φαρμακεία μεταξύ του Βόλου και των προαστίων του), τότε η χρήση ενός χωρικού δειγματοληπτικού υποβάθρου είναι αναγκαία.

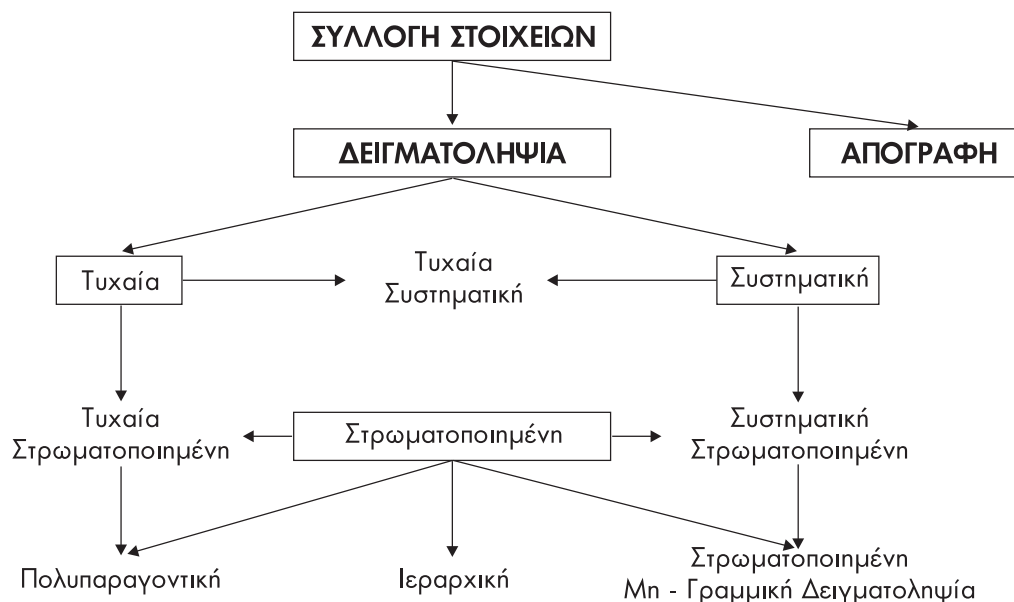
Από τα παραπάνω παραδείγματα φαίνεται καθαρά ότι το δειγματοληπτικό υπόβαθρο μπορεί να είναι χωρικό και μη χωρικό. Ένα **χωρικό δειγματοληπτικό υπόβαθρο** είναι εκείνο όπου η **θέση** είναι ένα βασικό τμήμα της μεταβλητότητας καθεμιάς από τις παρατηρήσεις και συνεπώς πρέπει απαραίτητα να αντιπροσωπεύει το συνολικό φάσμα της χωρικής μεταβλητότητας (*spatial variability*). Αν υπάρχει ανάγκη, για παράδειγμα, να ληφθεί ένα δείγμα από ένα χάρτη για να εξακριβωθεί το ποσοστό της γης που είναι καλυμμένο από δάση, πρέπει οπωσδήποτε όλα τα τμήματα του χάρτη να αντιπροσωπεύονται στο δειγματοληπτικό υπόβαθρο.

Επομένως, το δειγματοληπτικό υπόβαθρο μπορεί να είναι μη χωρικό και χωρικό, με αποτέλεσμα να υπάρχουν διάφοροι τρόποι επιλογής των μεμονωμένων παρατηρήσεων, που όμως όλες μαζί συνθέτουν το δείγμα. Όσον αφορά ένα μη χωρικό δειγματοληπτικό υπόβαθρο, αυτό συνήθως αποτελείται από έναν κατάλογο, από τα μέλη του οποίου επιλέγονται οι παρατηρήσεις που αναφέρονται σε αντικείμενα μιας συγκεκριμένης περιοχής (π.χ. ο κατάλογος των εστιατορίων της Αττικής, η θέση των οποίων μπορεί να είναι ή να μην είναι γνωστή). Στην περίπτωση ενός χωρικού δειγματοληπτικού υποβάθρου, όμως, πρέπει να καθοριστεί το είδος των παρατηρήσεων, δηλαδή να οριστεί η χωρική δειγματοληπτική μονάδα. Βασικά υπάρχουν τρεις τέτοιες μονάδες: **σημεία, γραμμές και επιφάνειες καννάβων** (*quad-*

rats), με αποτέλεσμα να υπάρχουν σημειακές και γραμμικές δειγματοληψίες καθώς και δειγματοληψίες επιφανειών καννάβων. Η χρησιμότητα και αποδοτικότητα καθεμιάς από αυτές τις δειγματοληψίες εξαρτάται από το αντικείμενο της έρευνας.

5.3. ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΗ ΔΙΑΔΙΚΑΣΙΑ

Τον καθορισμό του κατάλληλου δειγματοληπτικού υποβάθρου ακολουθεί η ουσιαστική δειγματοληψία, δηλαδή η διαδικασία επιλογής των παρατηρήσεων που θα αποτελέσουν το δείγμα. Η διαδικασία αυτή μπορεί να γίνει με δύο βασικούς τρόπους (*τυχαία και συστηματική*) και πολλούς σύνθετους με βάση κάποιας μορφής στρωματοποίηση (*π.χ. στρωματοποιημένη τυχαία, στρωματοποιημένη συστηματική μη γραμμική, ιεραρχική κ.λπ.*), όπως φαίνεται στο σχήμα 5.1.



Σχήμα 5.1: Είδη Δειγματοληψίας

5.3.1. Τυχαία Δειγματοληψία

Η τυχαία δειγματοληψία αποτελεί τη βάση κάθε στατιστικής επεξεργασίας στοιχείων και αναφέρεται σε μια διαδικασία που πληροί δύο βασικές στατιστικές συνθήκες: πρώτο τα μέλη ενός πληθυσμού έχουν ίσες πιθανότητες να επιλεγούν για το δείγμα και δεύτερο η επιλογή ενός μέλους του πληθυσμού για το δείγμα με κανένα

τρόπο δεν επηρεάζει την επιλογή ενός οποιουδήποτε άλλου. Δηλαδή οι παρατηρήσεις του δείγματος επιλέγονται **ανεξάρτητα** και με **ίσες πιθανότητες**.

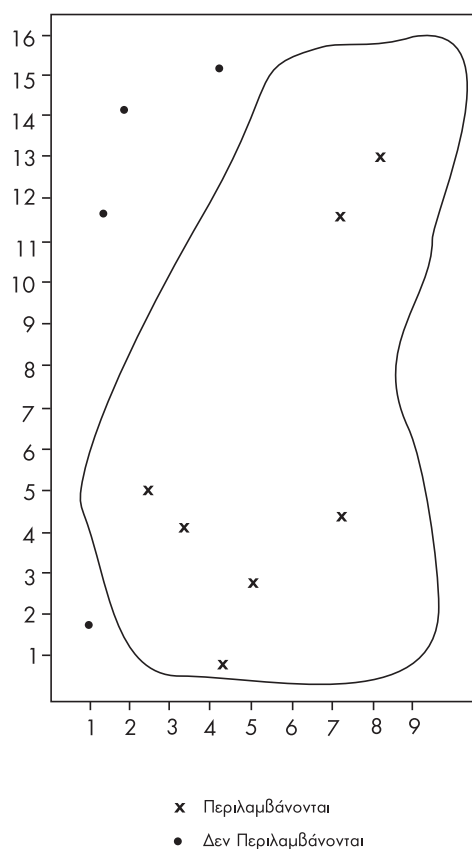
5.3.1.1. Δειγματοληψία από Κατάλογο: Όταν η χωρική διάσταση ενός πληθυσμού φαινομένων, αντικειμένων κ.λπ είναι μεν υπαρκτή αλλά όχι καθοριστική για την επίλυση του υπό εξέταση προβλήματος (*περιοχή ως σύνολο*), τότε η δειγματοληψία με τη χρήση ενός καταλόγου είναι όχι μόνο εφικτή αλλά και επιθυμητή. Για παράδειγμα, αν ο πληθυσμός που ενδιαφέρει είναι ο αριθμός των εργαζομένων στις 300 βιομηχανικές μονάδες του Θριάσιου Πεδίου, τότε το δειγματοληπτικό υπόβαθρο μπορεί να είναι ένας κατάλογος, κάθε μέλος (*βιομηχανική μονάδα*) του οποίου αντιπροσωπεύεται από έναν αριθμό (*π.χ. 1 έως 300*).

Η επιλογή ενός τυχαίου δείγματος από το δειγματοληπτικό αυτό υπόβαθρο επιτυγχάνεται με τη χρήση ενός πίνακα τυχαίων αριθμών, ο οποίος πληροί τις δύο συνθήκες τυχειότητας που αναφέρθηκαν προηγουμένως και είναι διαθέσιμος από πολλές πηγές, ως εξής: από έναν πίνακα των τυχαίων αριθμών (*Παράρτημα 5.1*) επιλέγεται μια τυχαία αφετηρία (*π.χ. η 3η τριάδα ψηφίων ακολουθώντας την πρώτη σειρά που στον πίνακα του Παραρτήματος 2.1 είναι ο αριθμός 080*) και στη συνέχεια επιλέγονται τριάδες ψηφίων με έναν άλλο επίσης τυχαίο τρόπο (*π.χ. κάθε 6η τριάδα ακολουθώντας αυτή τη φορά τη στήλη που στον πίνακα του Παραρτήματος 2.1 ο επόμενος αριθμός είναι 217*).

Κάθε τριάδα ψηφίων αντιστοιχεί σε ένα δείγμα που επιλέγεται (*στο συγκεκριμένο παράδειγμα οι μονάδες που επελέγησαν, αντιπροσωπεύονται από τους αριθμούς 80 και 217*) και η διαδικασία συνεχίζεται μέχρι να συμπληρωθεί ο αριθμός των δειγμάτων που έχει καθοριστεί.

Πρέπει να σημειωθεί ότι η επιλογή του αριθμού των ψηφίων εξαρτάται από το μέγεθος του πληθυσμού. Για παράδειγμα, αν ο αριθμός των μονάδων ήταν 3000 ή 33000, τότε οι επιλογές από τον πίνακα των τυχαίων αριθμών θα έπρεπε να είναι αντίστοιχα ομάδες τεσσάρων ή πέντε ψηφίων. Ακόμη, στην περίπτωση που επιλεγεί ένας άσχετος αριθμός (*π.χ. ο αριθμός 728 δεν αντιστοιχεί σε καμιά από τις 300 βιομηχανικές μονάδες*), ο αριθμός αυτός δεν λαμβάνεται υπόψη και η διαδικασία συνεχίζεται μέχρι να ολοκληρωθεί το δείγμα. Τέλος, η ίδια προσέγγιση ακολουθείται (*δεν λαμβάνεται υπόψη*), όταν ο ίδιος αριθμός επιλεγεί για δεύτερη ή περισσότερες φορές. Θεωρητικά, δεδομένης της αρχής της ανεξαρτησίας που ισχύει στον πίνακα των τυχαίων αριθμών, δεν τεκμηριώνεται η ενέργεια αυτή, πρακτικά όμως εξυπηρετεί χωρίς μεγάλα προβλήματα.

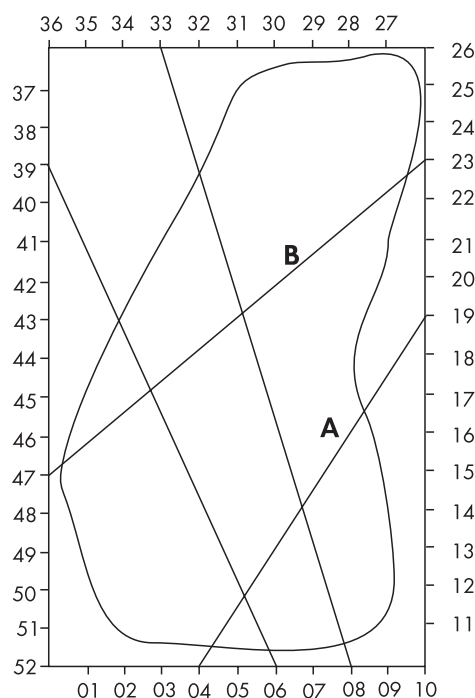
5.3.1.2. Σημειακή Δειγματοληψία: Η ίδια διαδικασία εφαρμόζεται όταν τα μέλη του πληθυσμού ορίζονται στο χάρτη ως σημεία, μια μορφή ιδιαίτερα δημοφιλής στη Γεωγραφία. Σε αντίθεση, όμως, με την προηγούμενη διαδικασία, αντί κάθε μέλος του πληθυσμού να αντιστοιχεί σε έναν αριθμό, αντιπροσωπεύεται από τη συντεταγμένη του (*τετμημένη και τεταγμένη*). Δηλαδή, χρησιμοποιώντας ένα σύστημα συντεταγμένων πάνω στο δειγματοληπτικό υπόβαθρο, δημιουργείται ένα σύνολο (*πληθυσμός*) από σημεία που ορίζονται από τις x και y συντεταγμένες τους. Στη συνέχεια με τη χρήση του πίνακα των τυχαίων αριθμών μπορεί να γίνει η επιλογή του τυχαίου δείγματος. Συνήθως, ως αφετηρία για τη χρήση του πίνακα επιλέγεται τυχαία ένας αριθμός και κατόπιν, σύμφωνα με ένα προκαθορισμένο τρόπο (*π.χ. κάθε νιοστός αριθμός διαγωνίως*), επιλέγονται οι υπόλοιποι αριθμοί που έχουν τον κατάλληλο αριθμό ψηφίων (*π.χ. τετραψήφιοι αριθμοί στην περίπτωση που οι άξονες μεταβάλλονται από 0 έως 100, όπου τα δύο πρώτα ψηφία τους ορίζουν την τετμημένη και τα δύο τελευταία την τεταγμένη του σημείου*). Όταν ο αριθμός που επελέγη, αντιστοιχεί στις συντεταγμένες ενός σημείου του δειγματοληπτικού υποβάθρου, τότε το σημείο αυτό συμπεριλαμβάνεται στο δείγμα και η επιλογή συνεχίζεται μέχρι να αποκτηθεί ένα δείγμα με όσα σημεία έχει υπολογιστεί ότι χρειάζονται. Εάν ο αριθμός που επιλέγεται από τον πίνακα των τυχαίων αριθμών αντιστοιχεί σε σημείο εκτός του δειγματοληπτικού υποβάθρου ή επιλέγεται για δεύτερη φορά, το σημείο αυτό δεν συμπεριλαμβάνεται στο δείγμα. Το σχήμα 5.2 δείχνει το αποτέλεσμα μιας τυχαίας δειγματοληψίας.



Σχήμα 5.2: Τυχαία Σημειακή Χωρική Δειγματοληψία

5.3.1.3. Γραμμική Δειγματοληψία: Με βάση την ίδια μεθοδολογία, που στηρίζεται στην αρχή της τυχαιότητας που εφαρμόστηκε προηγουμένως, οι πληροφορίες ενός χάρτη μπορούν να εκτιμηθούν με βάση τη γραμμική δειγματοληψία. Στην περίπτωση αυτή, όμως, αντί να υποθεθεί ότι ο χάρτης (το δειγματοληπτικό υπόβαθρο) αποτελείται από σημεία από τα οποία επιλέγεται το δείγμα, θεωρείται ότι η περιοχή μελέτης είναι μια συνεχής επιφάνεια και η αναζητούμενη πληροφορία επιλέγεται καταγράφοντάς την κατά μήκος μιας γραμμής.

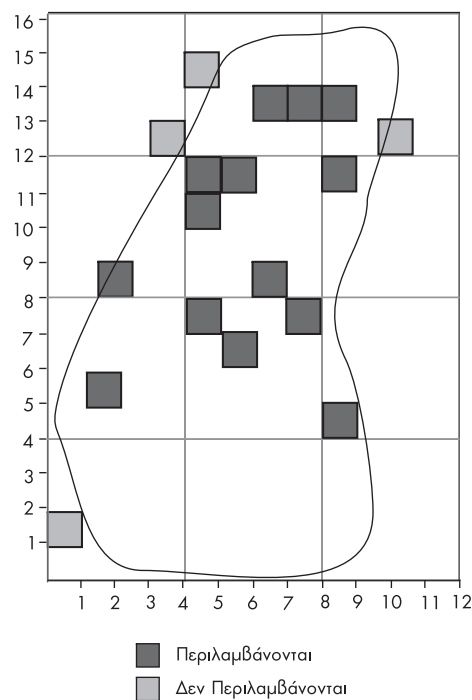
Υπάρχουν πολλοί τρόποι για τη δημιουργία μιας τυχάιας γραμμικής δειγματοληψίας. Ο πιο συνηθισμένος φαίνεται στο σχήμα 5.3, όπου τα περιθώρια του υποβάθρου βαθμονομούνται με τιμές που μπορούν να εκφραστούν με δύο (από 0 έως 99), με τρία (από 0 έως 999) ή περισσότερα ψηφία. Στη συνέχεια, με το γνωστό τρόπο επιλέγεται από τον πίνακα των τυχάιων αριθμών ένα σύνολο τετραψήφιων, οκταψήφιων κ.λπ αριθμών (ίσο με το μέγεθος του δείγματος) που καθορίζει την αρχή και το τέλος κάθε δειγματικής γραμμής. Για παράδειγμα, ο αριθμός 0419 του πίνακα ορίζει τη γραμμή Α (αρχή 04 και τέλος 19) και ο αριθμός 4723 τη γραμμή Β στο σχήμα 5.3. Στη συνέχεια καταγράφονται οι πληροφορίες που παρατηρούνται κατά μήκος της γραμμής (π.χ. το ποσοστό της γραμμής που «περνά» μέσα από διάφορες χρήσεις γης καθορίζει και την αναλογία τους).



Σχήμα 5.3: Τυχάια Γραμμική Χωρική Δειγματοληψία

5.3.1.4. Επιφανειακή (Κανναβική) Δειγματοληψία: Η δειγματοληψία αυτή αποτελεί συνέχεια των δύο προηγούμενων, καθώς ουσιαστικά διαφέρει μόνο ως προς τη δειγματοληπτική μονάδα. Πιο συγκεκριμένα, το δειγματοληπτικό υπόβαθρο θεωρείται

ότι αποτελείται από φατνία ενός καννάβου από τα οποία επιλέγεται τυχαία ένας μικρότερος αριθμός για να αποτελέσει το δείγμα. Η διαδικασία δειγματοληψίας είναι σχετικά απλή. Καταρχάς δημιουργείται ένα σύστημα συντεταγμένων που οριοθετεί το δειγματοληπτικό υπόβαθρο. Με το γνωστό τρόπο επιλέγονται από τον πίνακα των τυχαίων αριθμών ζεύγη, τετράδες κ.λπ ψηφίων τα οποία αντιστοιχούν στην κάτω αριστερή γωνία του φατνίου που επιλέγεται ως μέλος του δείγματος (σχήμα 5.4) και η επιλογή συνεχίζεται μέχρι να συμπληρωθεί το δείγμα. Και στην περίπτωση αυτή ισχύουν όσα ανα-φέρθηκαν στη σημειακή δειγματοληψία σχετικά με άσχετα φατνία (δηλαδή φατνία εκτός υποβάθρου) και φατνία που επιλέγονται περισσότερες φορές. Επιπλέον, όμως, δημιουργούνται προβλήματα, όταν το επιλεγμένο φατνίο δεν βρίσκεται εξ ολοκλήρου εντός του δειγματοληπτικού υποβάθρου. Η αντιμετώπιση όμως του προβλήματος αυτού ξεφεύγει από τους στόχους του παρόντος κεφαλαίου.



Σχήμα 5.4: Τυχαία Κανναβική
Χωρική Δειγματοληψία

Κατά τη διαδικασία αυτή σημαντικό ρόλο παίζει το σχήμα και το μέγεθος του φατνίου, στοιχεία τα οποία και καθορίζουν το αποτέλεσμα της δειγματοληψίας. Όσον αφορά το σχήμα, υπάρχει καθολική αποδοχή για το τετραγωνικό (κανναβικό) σχήμα. Όμως για το βέλτιστο μέγεθος του φατνίου έχουν διατυπωθεί διαφορετικές απόψεις (Haggett, 1965· Shimwell, 1971· Kershaw, 1965). Γενικά πάντως, πρέπει να προτιμάται ένας μεγάλος αριθμός μικρών φατνίων από ένα μικρό αριθμό μεγάλων φατνίων (Dixon and Leach, 1979).

Η τυχαία σημειακή, γραμμική και κανναβική μέθοδος έχουν τις ίδιες δυνατότητες στη δειγματοληψία χωρικών πληροφοριών. Απλώς η επιλογή μιας από αυτές τις μεθόδους εξαρτάται από τις συγκεκριμένες συνθήκες κάθε μελέτης. Για παράδειγμα, στις περισσότερες περιπτώσεις χαρτών που απεικονίζουν φαινόμενα, όπως γεωλογικοί

σχηματισμοί ή χρήσεις γης, έχει επιλεγεί η σημειακή. Εάν, όμως, το ενδιαφέρον εστιάζεται στη μέτρηση της πυκνότητας του υδρογραφικού δικτύου, αυτό μπορεί να επιτευχθεί με τον υπολογισμό του μήκους των ποταμών εντός μιας τυχαίας επιλεγείσας μικρής επιφάνειας (*φατνίο*). Κι αυτό γιατί με τη μικρή επιφάνεια που καταλαμβάνουν τα ποτάμια ή τα ρυάκια, μια τυχαία σημειακή δειγματοληψία είναι πιθανόν να μη δώσει αντιπροσωπευτικά αποτελέσματα. Επίσης, εάν ο στόχος της μελέτης είναι η εξακρίβωση της διαφοροποίησης των χρήσεων γης καθώς κάποιος απομακρύνεται από το κέντρο της πόλης, ένα γραμμικό δείγμα ακτινωτά από ένα κεντρικό σημείο σίγουρα θα προσφέρει ικανοποιητικά αποτελέσματα. Γενικά, στοιχεία όπως η καταλληλότητα, η ευκολία εφαρμογής καθώς και ο χρόνος που απαιτείται για την ολοκλήρωση του δείγματος αποτελούν καθοριστικούς παράγοντες επιλογής των επιμέρους δειγματοληψιών.

Τέλος, αξίζει να σημειωθεί ότι η τυχαία δειγματοληψία διαθέτει τα σπουδαία στατιστικά πλεονεκτήματα της ανεξαρτησίας και της ίσης πιθανότητας που βοηθούν στη στατιστική διαχείριση των αποτελεσμάτων της δειγματοληψίας. Πρακτικά, όμως, έχει το βασικό μειονέκτημα ότι οι διαδικασίες που αναφέρθηκαν προηγουμένως, δεν καταλήγουν πάντοτε σε μια ικανοποιητική χωρική κάλυψη του δειγματοληπτικού υποβάθρου. Συνήθως, η δειγματοληψία αποτελείται από δείγματα συγκεντρωμένα σε μια ή περισσότερες υποπεριοχές του υποβάθρου, με αποτέλεσμα σημαντικές χωρικές ιδιότητες του πληθυσμού να μην αντιπροσωπεύονται σε αυτό.

5.3.2. Συστηματική Δειγματοληψία

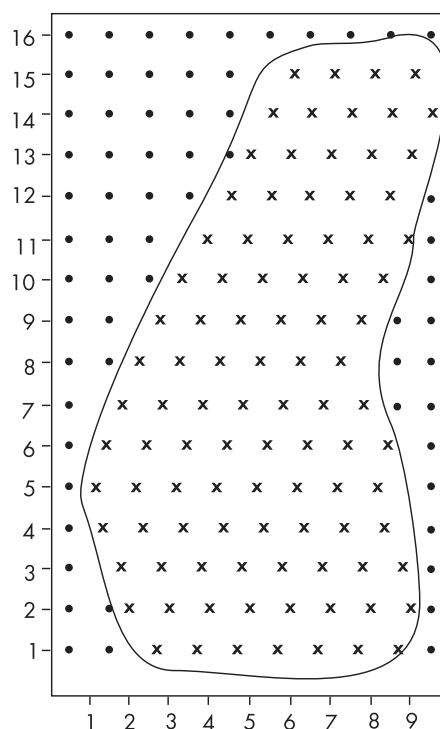
Προς αποφυγήν του μειονεκτήματος που διακρίνει την τυχαία δειγματοληψία, (*ελλιπής κάλυψη του δειγματοληπτικού υποβάθρου*) είναι δυνατόν να χρησιμοποιηθεί η συστηματική δειγματοληψία. Αυτή η διαδικασία απαιτεί, η επιλογή των επιμέρους παρατηρήσεων που αποτελούν το δείγμα, να γίνεται με ένα σταθερό και προκαθορισμένο (*συστηματικό*) τρόπο. Δηλαδή, αντί για τη χρήση τυχαίων αριθμών που ορίζουν τα μέλη του δείγματος σε οποιαδήποτε θέση του καταλόγου ή του χάρτη (*υποβάθρου*), το διάστημα μεταξύ των παρατηρήσεων να είναι σταθερό και καθορισμένο.

5.3.2.1. Δειγματοληψία από Κατάλογο: Στη δειγματοληψία αυτή χρησιμοποιείται για την επιλογή των παρατηρήσεων από τον κατάλογο, ένα σταθερό βήμα β , το οποίο ορίζεται ως το αντίστροφο της δειγματοληπτικής αναλογίας (n/N), δηλαδή του λόγου του δείγματος (n) προς τον πληθυσμό (N). Για παράδειγμα, σε έναν κατάλογο (*πληθυσμιακό υπόβαθρο*) 5000 αριθμών, εάν το δείγμα έχει καθοριστεί να είναι 50,

τότε το βήμα είναι το αντίστροφο του **50/5000**, δηλαδή **100**. Το συστηματικό δείγμα δημιουργείται ως εξής: Η αφετηρία επιλέγεται τυχαία και είναι συνήθως ένας αριθμός μεταξύ **1** και **β** , έτσι ώστε να διασφαλίζεται ότι όλοι οι αριθμοί του καταλόγου, πριν επιλεγεί ο πρώτος, έχουν πιθανότητα να επιλεγούν. Ο πρώτος αριθμός είναι και ο μοναδικός που επιλέγεται τυχαία. Στη συνέχεια, η διαδικασία επιλογής των άλλων μελών του δείγματος είναι απλή αλλά προκαθορισμένη, αφού κάθε φορά επιλέγεται ο **β** επόμενος αριθμός.

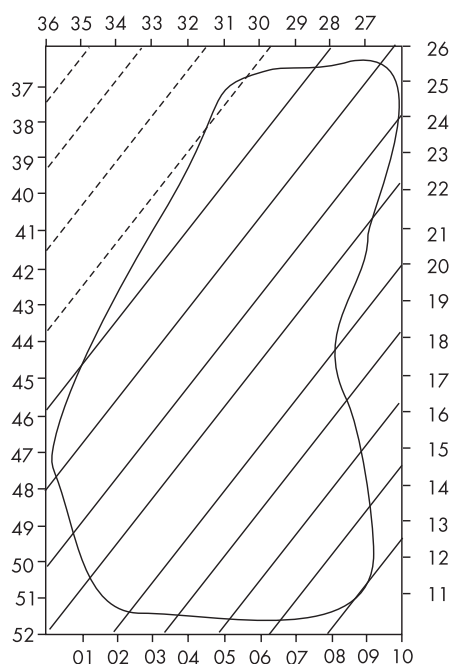
Σε μια τέτοια συστηματική δειγματοληψία είναι εύκολο να ελεγχθεί η σωστή εφαρμογή της, εξαιτίας της κανονικότητας του βήματος επιλογής των παρατηρήσεων. Επιπλέον, το δείγμα προέρχεται από κάθε μέρος του δειγματοληπτικού υποβάθρου (κατάλογος) και με τις σωστές αναλογίες.

5.3.2.2. Σημειακή Δειγματοληψία: Η ίδια διαδικασία ακολουθείται, όταν το δειγματοληπτικό υπόβαθρο είναι ένας χάρτης και τα μέλη του πληθυσμού θεωρούνται ότι απεικονίζονται με σημεία. Πιο συγκεκριμένα, η δειγματοληψία αρχίζει από ένα τυχαία επιλεγμένο σημείο ακριβώς με τον ίδιο τρόπο που επιλέγονται τα σημεία στην τυχαία δειγματοληψία και κατόπιν με ένα προκαθορισμένο σχέδιο (κάθε σημείο σε ίση απόσταση, κατά τους άξονες x και y , από το άλλο) επιλέγονται και τα υπόλοιπα σημεία του δείγματος. Το σχήμα 5.5 δείχνει το αποτέλεσμα μιας σημειακής συστηματικής δειγματοληψίας.



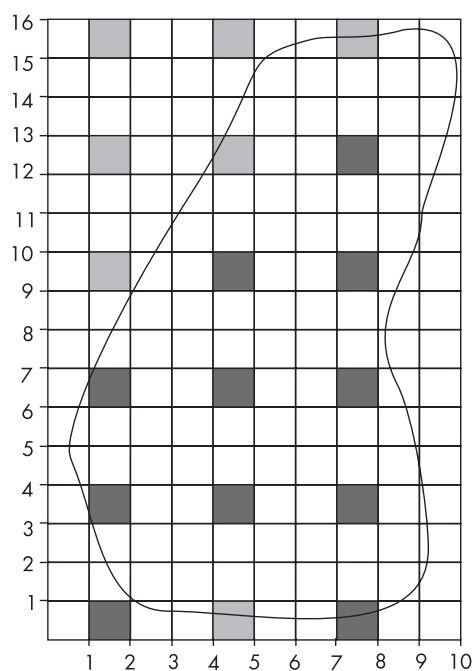
Σχήμα 5.5: Συστηματική Σημειακή
Χωρική Δειγματοληψία

5.3.2.3. Γραμμική Δειγματοληψία: Με βάση το ίδιο σκεπτικό επιτυγχάνεται και η γραμμική δειγματοληψία. Στη συγκεκριμένη περίπτωση επιλέγεται με τυχαίο τρόπο η πρώτη γραμμή (βλέπε τυχαία γραμμική δειγματοληψία) και το δειγματοληπτικό υπόβαθρο καλύπτεται με μια σειρά από παράλληλες προς την αρχική γραμμές σε μια συγκεκριμένη απόσταση μετρούμενη στον άξονα της αρχής της πρώτης γραμμής, όπως φαίνεται στο σχήμα 5.6.



Σχήμα 5.6: Συστηματική Γραμμική Χωρική Δειγματοληψία

5.3.2.4. Κανναβική Δειγματοληψία: Στην περίπτωση που το ενδιαφέρον εστιάζεται σε χωρικά φαινόμενα, όπου η επιφάνεια αποτελεί συστατικό στοιχείο (π.χ. η μέση πυκνότητα κάποιων μορφής κτηρίων στο κέντρο της πόλης), μπορεί να χρησιμοποιηθεί η κανναβική δειγματοληψία, η οποία αποτελεί απλώς μια συνέχεια των προηγούμενων, αλλά με δειγματική μονάδα την τετραγωνική επιφάνεια. Στη δειγματοληψία αυτή, η πρώτη δειγματική επιφάνεια επιλέγεται με τον ίδιο



Σχήμα 5.7: Συστηματική Κανναβική Χωρική Δειγματοληψία

τρόπο που ακολουθείται και στην τυχαία επιφανειακή δειγματοληψία και στη συνέχεια τα επόμενα δείγματα επιλέγονται με τέτοιο τρόπο ώστε η κάτω αριστερή γωνία κάθε φατνίου να βρίσκεται σε μια προκαθορισμένη απόσταση, κατά τους άξονες x και y , από το προηγούμενο (σχήμα 5.7). Η δειγματοληψία αυτή είναι ιδιαίτερα χρήσιμη, όταν απαιτείται δειγματοληψία στην ύπαιθρο.

Η συστηματική δειγματοληψία έχει δύο βασικά μειονεκτήματα. Πρώτο, το αρχικό σημείο που επιλέγεται, καθορίζει και την επιλογή των υπόλοιπων σημείων. Επομένως, όλα τα σημεία του δειγματοληπτικού υποβάθρου δεν έχουν την ίδια πιθανότητα επιλογής μετά τον ορισμό του πρώτου σημείου του δείγματος. Αυτό σημαίνει ότι όλα τα τμήματα της περιοχής που μελετάται, δεν έχουν ίση ευκαιρία για να συμπεριληφθούν στο δείγμα, με αποτέλεσμα από τη σκοπιά της στατιστικής (*δηλαδή την εκτίμηση των πληθυσμιακών παραμέτρων*) η διαδικασία αυτή να δημιουργεί σοβαρά ερωτηματικά. Δεύτερο, με τη συστηματική δειγματοληψία, αν υπάρχουν περιοδικότητες στο δειγματοληπτικό υπόβαθρο, η συλλογή σημείων σε κανονικά διαστήματα θα έχει ως αποτέλεσμα όλα τα σημεία του δείγματος να εκφράζουν κάθε φορά το ίδιο χαρακτηριστικό της περιοδικότητας, δίνοντας έτσι μια μεροληπτική εικόνα των χωρικών διαφοροποιήσεων των φαινομένων που μελετώνται (*Berry and Baker, 1968*). Από την άλλη μεριά όμως, είναι κατά πολύ γρηγορότερο και απλούστερο να ληφθεί ένα συστηματικό δείγμα από ότι ένα τυχαίο, αφού δεν απαιτείται η συνεχής χρήση του πίνακα των τυχαίων αριθμών. Επιπλέον, το συστηματικό δείγμα δίνει περισσότερο ομοιόμορφη κάλυψη του δειγματοληπτικού υποβάθρου. Δηλαδή, σε πολλές περιπτώσεις η μεροληπτικότητα του συστηματικού δείγματος είναι τόσο μικρή, ώστε τα οφέλη σε χρόνο και άλλα διαθέσιμα που προκύπτουν από τη χρήση τους να αντισταθμίζουν κατά πολύ τα προβλήματα από αυτά τα σφάλματα.

5.3.2.5. Συστηματικές Τυχαίες Δειγματοληψίες: Μερικά από τα προβλήματα των συστηματικών δειγμάτων μπορούν να ελαχιστοποιηθούν με το συνδυασμό της συστηματικής δειγματοληψίας με μια μεγαλύτερη συμμετοχή της τυχαίας διαδικασίας, που μπορεί να επιτευχθεί με μια σειρά από τρόπους:

- **Πολλαπλή Επιλογή Αφετηρίας:** Στην περίπτωση που υπάρχουν υπόνοιες για κάποιας μορφής περιοδικότητα, μια απλή μέθοδος αντιμετώπισής της είναι η εξής: Αντί για την επιλογή μιας μόνο τυχαίας αφετηρίας, επιλέγονται περισσότερες και για κάθε τυχαία αφετηρία ακολουθείται η γνωστή διαδικασία. Για παράδειγμα, στην περίπτωση δειγματοληψίας από κατάλογο με $\beta=10$, αν επιλεγούν οι τυχαίοι αριθμοί 5, 28, 267,

396 και 583, τότε το δείγμα θα αποτελείται από τους αριθμούς 5, 15, 25 και στη συνέχεια 28, 38, 48 ..., 267, 277, 287 ..., 396, 406, 416 ..., 583, 593 ...

- **Τυχαίο Διάστημα:** Κατά τη διαδικασία αυτή κάθε δείγμα επιλέγεται τυχαία, αλλά η επιλογή του αποκλείει την επιλογή οποιουδήποτε άλλου δείγματος που βρίσκεται εντός μιας συγκεκριμένης απόστασης. Για παράδειγμα, σε μια σημειακή δειγματοληψία η επιλογή ενός σημείου αποκλείει την επιλογή όλων των σημείων που βρίσκονται εντός ενός κύκλου με προκαθορισμένη ακτίνα α . Καθώς με την τυχαία επιλογή οι αποστάσεις μεταξύ των σημείων πλησιάζουν την ακτίνα α , τότε η δειγματοληψία πλησιάζει τη συστηματική. Η διαδικασία αυτή έχει το πλεονέκτημα της πλήρους κάλυψης του υποβάθρου αλλά, σε αντίθεση με τις υπόλοιπες συστηματικές προσεγγίσεις, δεν παρουσιάζει το μειονέκτημα της τυχαίας περάτωσης.
- **Τυχαία Μεταβαλλόμενο Βήμα:** Στη δειγματοληψία αυτή το βήμα για τη συστηματική επιλογή δεν είναι σταθερό, αλλά κάθε φορά μεταβάλλεται κατά ένα μικρό τυχαίο αριθμό λ που προστίθεται ή αφαιρείται. Επομένως, τα μέλη του δείγματος που επιλέγονται, απέχουν κάθε φορά $\pm\lambda$. Για παράδειγμα, το διάστημα μεταξύ των αριθμών που επιλέγονται σε έναν κατάλογο είναι $\beta \pm \lambda$ (με εκκίνηση μεταξύ του I και K), ενώ σε σημειακό υπόβαθρο η απόσταση κατά τους άξονες των X και Y θα είναι $\alpha_x \pm \lambda$ και $\alpha_y \pm \lambda$. Εάν η τιμή λ θα προστεθεί ή θα αφαιρεθεί, καθορίζεται τυχαία. Ένας τρόπος, για παράδειγμα, είναι να επιλέγεται από τον πίνακα τυχαίων αριθμών ένας μονοψήφιος αριθμός και αν είναι μονός, να προστίθεται ενώ αν είναι ζυγός, να αφαιρείται το λ .

5.3.3. Στρωματοποιημένη Δειγματοληψία

Μια περαιτέρω βελτίωση των προηγούμενων διαδικασιών μπορεί να επέλθει μέσα από μια βασική σύνθετη δειγματοληψία που είναι η χρήση των στρωματοποιημένων δειγμάτων, όπου ο πληθυσμός διαιρείται σε μικρότερες ομάδες που δειγματοληπτούνται χωριστά. Αυτή η διαδικασία είναι πολύ σημαντική, όταν ο πληθυσμός αποτελείται από διακεκριμένες υποομάδες που χαρακτηρίζονται από διαφορετικό μέγεθος ή χαρακτηριστικά. Για παράδειγμα, για να καθοριστεί σε μια δειγματοληψία η έκταση που καταλαμβάνουν τα καταστήματα λιανικής πώλησης, είναι πολύ λογικό να περιφερειοποιηθεί η πόλη με βάση την απόσταση από το κέντρο και μετά να ληφθούν δείγματα σε κάθε ζώνη. Και αυτό επειδή το ποσοστό γης που καταλαμβάνεται από τα καταστήματα λιανικής πώλησης ελαττώνεται, όσο κάποιος απομακρύνεται από το

κέντρο και επομένως θα πρέπει το μέγεθος του δείγματος σε κάθε ζώνη να ελαττώνεται ανάλογα.

Δύο επιπλέον καταστάσεις που επιβάλλουν τη χρήση στρωματοποιημένης δειγματοληψίας είναι: πρώτον, η σύγκριση κάποιου χωρικού χαρακτηριστικού δύο (ή περισσότερων) περιοχών. Για παράδειγμα, αν υπάρχει ανάγκη να εξεταστούν αγροτεμάχια στο νομό Μαγνησίας με την πρόθεση να συγκριθεί το μέγεθος των αγροτεμαχίων στα παράλια και στο εσωτερικό του νομού, θα ήταν λογικό να στρωματοποιηθεί ο νομός σε παραλίες και εσωτερικές περιοχές. Δεύτερον, η σύγκριση δύο μεταβλητών, όταν δε διαφοροποιούνται ομοιόμορφα στο χώρο του δειγματοληπτικού υποβάθρου. Για παράδειγμα, αν εξετάζεται η σχέση μεταξύ του είδους της καλλιέργειας και του υψομέτρου στο νομό Μαγνησίας, θα ήταν λογικό να υπάρχει αντιπροσώπευση όλων των υψομετρικών ζωνών. Ένα απλό τυχαίο δείγμα είναι σίγουρο πως δεν θα έδινε μια τέτοια αντιπροσώπευση, γιατί η μεγαλύτερη επιφάνεια του νομού Μαγνησίας ανήκει σε χαμηλά υψόμετρα (π.χ. μέχρι 100μ. υψόμετρο).

Πρέπει να σημειωθεί ότι υπάρχουν δύο βασικά πλεονεκτήματα με τη χρήση της στρωματοποιημένης δειγματοληψίας. Πρώτον, με την προσέγγιση αυτή εξασφαλίζεται ότι κάθε τμήμα του υπό εξέταση πληθυσμού συμπεριλαμβάνεται στο δείγμα, ώστε να είναι αντιπροσωπευτικό. Ένα δεύτερο πλεονέκτημα είναι ότι η στρωματοποίηση αυξάνει σημαντικά την ποιότητα των εκτιμήσεων για τον πληθυσμό, που σε τελική ανάλυση είναι και ο αποκλειστικός σκοπός κάθε δειγματοληψίας. Για παράδειγμα, αν θεωρηθεί ότι οι διάφορες χρήσεις γης καθορίζουν συγκεκριμένα χαρακτηριστικά των κοινωνικών ομάδων, τότε, όσο πιο διαφορετικές είναι οι κοινωνικές ομάδες αυτές, τόσο περισσότερο θα βελτιώνονται οι πληθυσμιακές εκτιμήσεις από τη δημιουργία στρωμάτων στο δείγμα.

Όσον αφορά το μέγεθος του δείγματος για κάθε «στρώμα» του δειγματοληπτικού υποβάθρου, επειδή δεν υπάρχει κανένας λόγος να είναι το ίδιο -το αντίθετο μάλιστα-, η πιο απλή προσέγγιση και αυτή που συνήθως χρησιμοποιείται είναι η αναλογική κατανομή του συνολικού δείγματος με βάση τον αριθμό κάθε υποομάδας στον κατάλογο ή την επιφάνεια κάθε υποπεριοχής στο χάρτη. Συχνά, όμως, μια τέτοια προσέγγιση δεν είναι ικανοποιητική, όπως για παράδειγμα στην περίπτωση που η μεταβλητότητα μεταξύ των «στρωμάτων» διαφέρει, με αποτέλεσμα κάθε «στρώμα» να χρειάζεται ένα «βάρος» που να τροποποιεί την απλή αναλογική εκπροσώπηση (για περισσότερες πληροφορίες καλές πηγές είναι οι: Dixon and Leach, 1979 και Stopher and Meyburg, 1979).

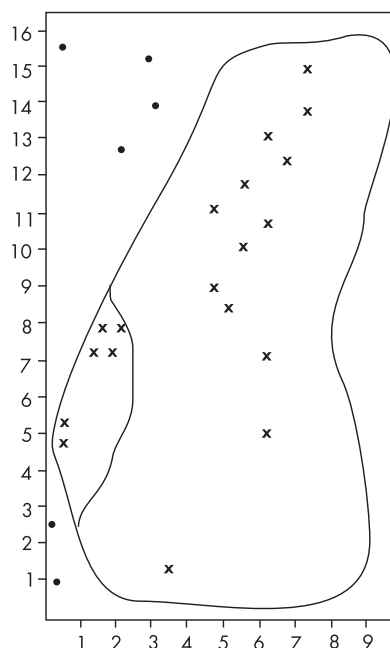
5.3.3.1. Στρωματοποίηση σε Κατάλογο: Η στρωματοποίηση στη συγκεκριμένη περίπτωση είναι φανερό ότι είναι ιδιαίτερα εύκολη. Για παράδειγμα, αν υπάρχει ανάγκη να ληφθεί δείγμα για να διαπιστωθεί ο τρόπος που αντιλαμβάνονται την απόσταση οι άνδρες και οι γυναίκες, δημότες της κοινότητας Πικερμίου Αττικής, μια τέτοια στρωματοποίηση είναι εφικτή και εύκολη. Απλώς λαμβάνονται ίσα δείγματα ανδρών και γυναικών από τον κατάλογο των δημοτών Πικερμίου. Με τον ίδιο τρόπο, αν το ενδιαφέρον είναι στο μέσο μέγεθος των καταστημάτων του Βόλου, η στρωματοποίηση σε κατηγορίες καταστημάτων οδηγεί στη λήψη δειγμάτων σε αναλογία με τον αριθμό των καταστημάτων τροφίμων, εμπορικών, ειδών υγιεινής κ.λπ.

Στα παραπάνω παραδείγματα αλλά και σε άλλα παρόμοια τα χαρακτηριστικά που διαφοροποιούν τα «στρώματα» είναι γνωστά και ευκρινή. Δε συμβαίνει το ίδιο, όμως, σε όλες τις περιπτώσεις και ειδικότερα σ' εκείνες που το δειγματοληπτικό υπόβαθρο δεν παρέχει πληροφορίες για τη μεταβλητή στην οποία εστιάζεται η μελέτη και καθορίζει τη στρωματοποίηση. Επομένως, πρέπει να γίνει χρήση άλλων πηγών, μια διαδικασία αρκετά δύσκολη και χρονοβόρα αλλά πάντοτε εφικτή.

Στη Γεωγραφία και γενικά στις επιστήμες που «θεραπεύουν» το χώρο, η στρωματοποίηση είναι ιδιαίτερα χρήσιμη κυρίως κατά τη δειγματοληψία σημείων τόσο στο χάρτη (στο γραφείο) όσο και στην ύπαιθρο. Στις περιπτώσεις αυτές υπάρχουν δύο προσεγγίσεις: η πρώτη αναφέρεται σε δειγματοληπτικό υπόβαθρο στο οποίο μπορούν να διακριθούν σαφώς φυσικές υποπεριοχές γνωστής έκτασης, ενώ στη δεύτερη απαιτείται η τεχνική υποδιαίρεση του υποβάθρου με τη βοήθεια ενός καννάβου.

5.3.3.2. «Φυσική» Στρωματοποίηση: Η χωρική αυτή δειγματοληψία αναφέρεται στην ανεξάρτητη λήψη δειγμάτων σε κάθε «στρώμα» της περιοχής μελέτης. Το μέγεθος του δείγματος σε κάθε υποπεριοχή είναι ανάλογο της επιφάνειάς της, η οποία λόγω της φυσικής ύπαρξής της είναι γνωστή, ενώ το είδος (βασική μορφή) της δειγματοληψίας εξαρτάται από το φαινόμενο που μελετάται. Επομένως, δύο είναι οι δειγματοληπτικές προσεγγίσεις για σημειακά στρωματοποιημένα φαινόμενα.

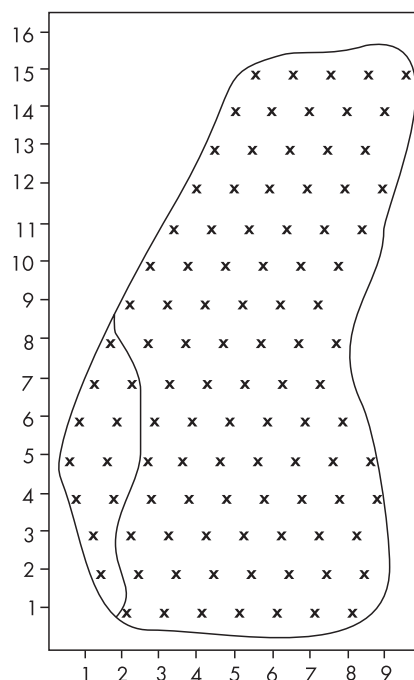
- **Στρωματοποιημένη Τυχαία Δειγματοληψία:** Στην περίπτωση αυτή η δειγματοληψία επιτυγχάνεται με τη



Σχήμα 5.8: Στρωματοποιημένη Τυχαία Δειγματοληψία

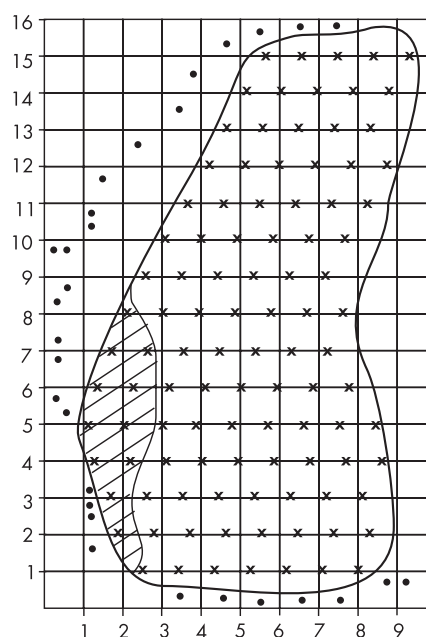
στρωματοποίηση του υποβάθρου σε υποπεριοχές και στη συνέχεια σε **καθεμιά** από αυτές γίνεται τυχαία επιλογή δειγμάτων με το γνωστό τρόπο. Στο παράδειγμα που αναφέρθηκε για τις καλλιέργειες του Νομού Μαγνησίας, μια τέτοια δειγματοληψία μπορεί να επιτευχθεί με τη στρωματοποίηση του νομού σε υψομετρικές ζώνες (π.χ. 0-100μ, 100-150μ, 150-200μ. κλπ) και την επιλογή με τυχαία δειγματοληψία ενός ορισμένου αριθμού σημείων σε κάθε ζώνη ανάλογα με το μέγεθός της. Η μορφή της δειγματοληψίας αυτής φαίνεται στο σχήμα 5.8.

- Στρωματοποιημένη Συστηματική Δειγματοληψία:** Η δειγματοληψία αυτή διαφέρει από την προηγούμενη μόνο ως προς την επιλογή του δείγματος στις υποπεριοχές που είναι συστηματική και όχι τυχαία (σχήμα 5.9). Για παράδειγμα, στην περίπτωση του μεγέθους των αγροτεμαχίων του νομού Μαγνησίας, τη στρωματοποίηση σε παραλία και εσωτερικές περιοχές ακολουθούν δύο συστηματικά δείγματα (ένα για κάθε περιοχή). Η επιλογή του συστηματικού δείγματος είναι αποτέλεσμα της ανάγκης πλήρους κάλυψης των υποπεριοχών που χαρακτηρίζονται από μεγάλη μεταβλητότητα (οι εσωτερικές περιοχές περιλαμβάνουν μικρού μεγέθους αγροκτήματα στο Πήλιο και μεγάλο στον Αλμυρό).



Σχήμα 5.9: Στρωματοποιημένη Συστηματική Δειγματοληψία

5.3.3.3. Κανναβική Στρωματοποίηση: Για την αποφυγή προβλημάτων, όπως αυτό που αναφέρθηκε προηγουμένως, δηλαδή της μεγάλης μεταβλητότητας στην περιοχή μελέτης και γενικά της ανάγκης συνδυασμού στρωματοποίησης, συστηματικότητας και τυχειότητας, αλλά κυρίως όταν απαιτείται εργασία υπαίθρου, όπου η χρήση

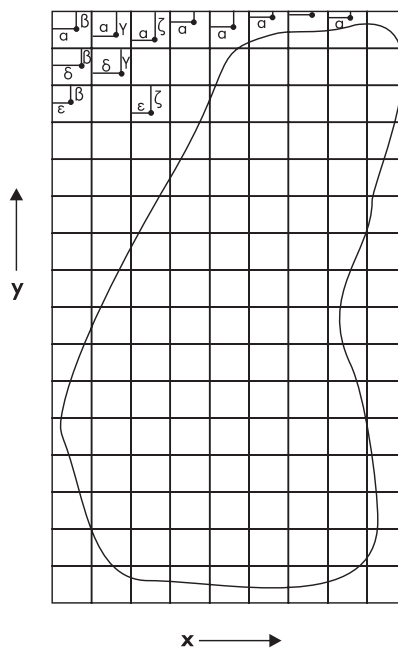


Σχήμα 5.10: Κανναβική Τυχαία Στρωματοποιημένη Δειγματοληψία

των χωρικών δειγματοληψιών που εφαρμόζονται στο γραφείο σε χάρτες δεν είναι ιδιαίτερα εύκολη και εφικτή, τότε προτιμάται η κανναβική στρωματοποίηση. Πιο συγκεκριμένα, η περιοχή μελέτης επικαλύπτεται από ένα κάνναβο με φατνία ίσου μεγέθους και για κάθε υποπεριοχή 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 (σχήμα 5.10).

5.3.2.4. Στρωματοποιημένη Συστηματική μη Γραμμική Δειγματοληψία: Η στρωματοποιημένη συστηματική μη γραμμική δειγματοληψία αποτελεί μια άλλη σύνθεση των τριών τρόπων δειγματοληψίας και προτάθηκε από τους Berry and Baker (1968). Συγκεκριμένα, η δειγματοληψία αυτή συνδυάζει τα πλεονεκτήματα της τυχαίας επιλογής και της στρωματοποίησης μαζί με τα χρήσιμα χαρακτηριστικά των συστηματικών δειγμάτων.

Η δειγματοληψία ενός στρωματοποιημένου συστηματικού μη γραμμικού δείγματος ακολουθεί μια σχετικά σύνθετη διαδικασία. Στην αρχή το δειγματοληπτικό υπόβαθρο διαιρείται με βάση ένα σύστημα συντεταγμένων σε κάνναβο με φατνία ίσου μεγέθους. Το πρώτο σημείο επιλέγεται τυχαία στο πρώτο γωνιακό φατνίο (συντεταγμένες α , β). Οι συντεταγμένες του σημείου αυτού καθορίζουν την επιλογή όλων των υπολοίπων σημείων. Συγκεκριμένα, η τετμημένη του σημείου αυτού α κρατιέται σταθερή για όλα τα σημεία που επιλέγονται στα φατνία της πρώτης σειράς (ένα σημείο σε κάθε φατνίο), ενώ η τεταγμένη τους επιλέγεται τυχαία από τους πίνακες των τυχαίων αριθμών. Παρόμοια, η τεταγμένη του πρώτου σημείου β αποτελεί την τεταγμένη όλων των σημείων των φατνίων της πρώτης στήλης που η τετμημένη τους επιλέγεται τυχαία (βλέπε σχήμα 5.11). Η διαδικασία συνεχίζεται με την επιλογή των συντεταγμένων ενός δεύτερου σημείου που ανήκει στο φατνίο της δεύτερης σειράς και της δεύτερης στήλης. Ορίζεται να έχει το σημείο αυτό ως τετμημένη, την τετμημένη δ του σημείου της πρώτης στήλης και της δεύτερης σειράς και ως τεταγμένη, την τεταγμένη γ του σημείου της δεύτερης στήλης και της πρώτης σειράς που ορίστηκαν τυχαία. Ανάλογα ορίζεται και το σημείο της τρίτης στήλης, της τρίτης σειράς με συντεταγμένες ϵ και ζ και η

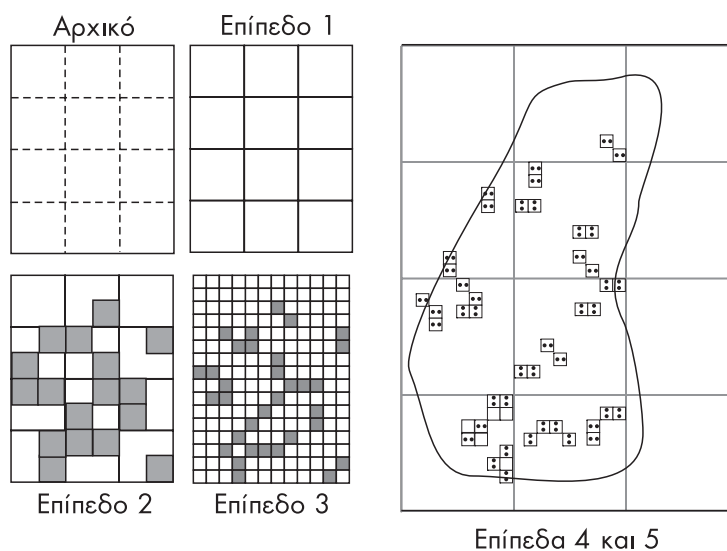


Σχήμα 5.11: Στρωματοποιημένη Συστηματική μη Γραμμική Δειγματοληψία

διαδικασία συνεχίζεται μέχρι να επιλεγούν σημεία για κάθε φατνίο του δειγματοληπτικού υποβάθρου με τέτοιο τρόπο, ώστε να είναι σταθερή η τετμημένη σε κάθε σειρά και η τεταγμένη σε κάθε στήλη.

5.3.5. Ιεραρχική Δειγματοληψία

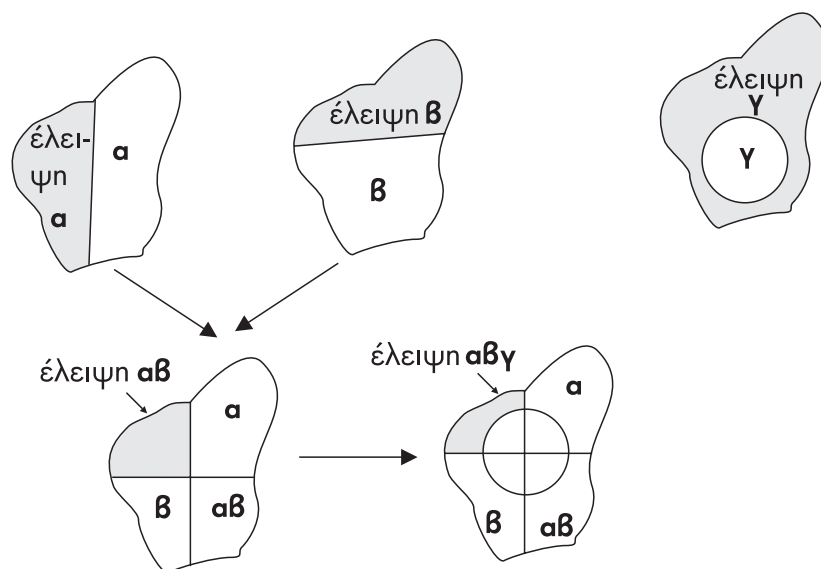
Σε ορισμένες ερευνητικές προσπάθειες είναι επιθυμητό να ληφθούν δείγματα με αυξανόμενη λεπτομέρεια σε διάφορα επίπεδα. Έτσι μπορεί, για παράδειγμα, να επιλεγεί τυχαία ένας ή περισσότεροι νομοί, κατόπιν μέσα σε αυτούς τους νομούς να επιλεγεί τυχαία ένας αριθμός επαρχιών και τέλος από αυτές τις επαρχίες να επιλεγούν τυχαία μερικά αγροτεμάχια (*αν ερευνάται το μέγεθος των αγροτεμαχίων*). Δηλαδή, το δειγματοληπτικό υπόβαθρο διαιρείται σε μια ιεραρχία δειγματοληπτικών επιπέδων, τα οποία μπορεί να έχουν ακανόνιστο ή κανναβικό σχήμα και τα οποία εμπεριέχονται το ένα μέσα στο άλλο. Ο δειγματοληπτικός αυτός τρόπος έχει τα εξής πλεονεκτήματα: πρώτον, μειώνει σημαντικά τα έξοδα της επιτόπιας έρευνας (*αν απαιτείται επιτόπια έρευνα*) και, δεύτερον, είναι ιδιαίτερα επιτυχημένος στη διερεύνηση διαφοροποιήσεων σε διάφορες χωρικές κλίμακες. Έχει όμως το μειονέκτημα της σχετικά μεγάλης στατιστικής αναξιοπιστίας. Στο σχήμα 5.12 παρουσιάζεται μια κανναβική ιεραρχική δειγματοληψία πέντε επιπέδων που μπορεί να εφαρμοστεί τόσο στην περίπτωση επιφανειών (*τα τέσσερα πρώτα επίπεδα*) όσο και στην περίπτωση σημείων (*πέμπτο επίπεδο*).



Σχήμα 5.12: Κανναβική Ιεραρχική Δειγματοληψία

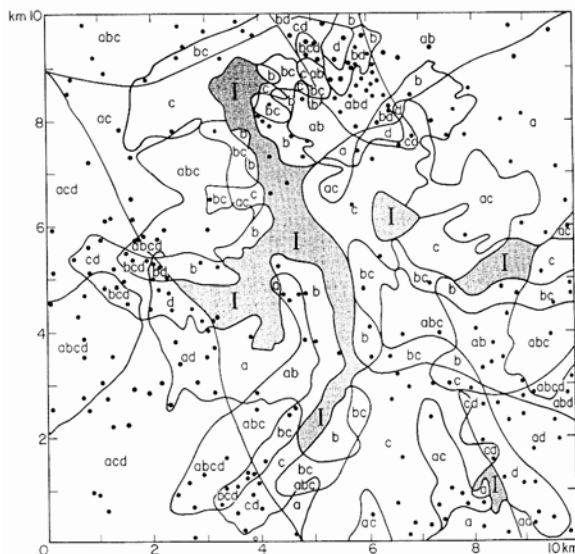
5.3.6. Πολυπαραγοντική Δειγματοληψία

Σε περιπτώσεις που το ενδιαφέρον εστιάζεται στο συνδυασμό δύο ή περισσότερων παραγόντων οι οποίοι θεωρούνται ότι επηρεάζουν τη χωρική κατανομή ενός φαινομένου (π.χ. χρήσεις γης), τότε χρησιμοποιείται η πολυπαραγοντική δειγματοληψία. Στη συγκεκριμένη δειγματοληψία η περιοχή μελέτης χωρίζεται σε υποπεριοχές οι οποίες εκφράζουν κάθε παράγοντα και η επικάλυψή τους εκφράζει το συνδυασμό τους, που με τη σειρά του δημιουργεί ένα σύνολο από πολυπαραγοντικές περιοχές, όπως παρουσιάζεται σχηματικά στο σχήμα 5.13.



Σχήμα 5.13: Δημιουργία Πολυπαραγοντικής Δειγματοληψίας

Σε κάθε υποπεριοχή επιλέγονται τυχαία σημεία με το γνωστό τρόπο. Τα αποτελέσματα της δειγματοληψίας αυτής με τη χρήση της ανάλυσης διασποράς μπορούν να βοηθήσουν στον υπολογισμό της επίδρασης κάθε παράγοντα και την αλληλεπίδρασή τους. Μια καλή εφαρμογή δίνεται από τον Haggett (1964) και φαίνεται στο σχήμα 5.14.



Σχήμα 5.14: Παράδειγμα Πολυπαραγοντικής Δειγματοληψίας

Πηγή: Haggett, Regroul and Local Campanb, 1964

5.3.7. Χωρο-Χρονική Δειγματοληψία

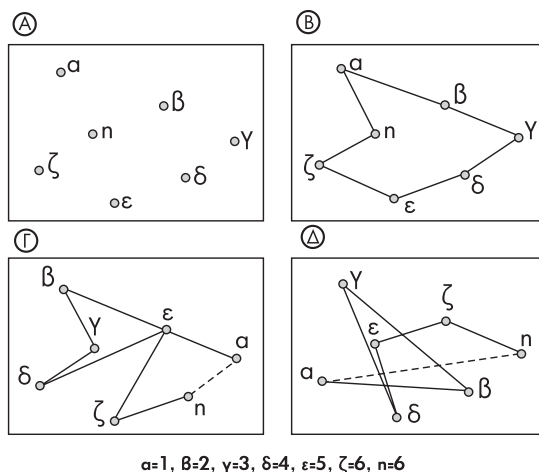
Όλοι οι τρόποι λήψης δειγμάτων που προαναφέρθηκαν είναι πραγματικά χρήσιμοι, όταν το ενδιαφέρον εστιάζεται στη χωρική διαφοροποίηση φαινομένων, χαρακτηριστικών κ.λπ. Το ερώτημα που τίθεται, όμως, είναι: Υπάρχει τρόπος εξέτασης χαρακτηριστικών που εκτός των άλλων μεταβάλλονται και με το χρόνο; Τυπικό παράδειγμα αποτελεί το χωρικό πρότυπο των κυκλοφοριακών φόρτων σε μια πόλη. Οι κυκλοφοριακοί φόρτοι διαφέρουν από μια διασταύρωση σε μια άλλη (σημεία), αλλά επίσης μεταβάλλονται και από ώρα σε ώρα κατά τη διάρκεια του 24ώρου. Μια πλήρης μελέτη θα απαιτούσε ένα πυκνό δίκτυο σημείων για την καταγραφή των φόρτων στα οποία επιπλέον θα ελαμβάνοντο μετρήσεις συνεχώς. Επειδή, όπως έχει ειπωθεί επανειλημμένως, ένα σωστό δείγμα μπορεί να μετρήσει εξίσου καλά τους φόρτους αυτούς με σαφώς μικρότερες δαπάνες και προσπάθειες, το πρόβλημα μετατίθεται στη λήψη ενός ικανοποιητικού χωρο-χρονικού δείγματος.

Μια συχνά χρησιμοποιούμενη δειγματοληψία έχει προταθεί από τον Haggett (1975) και φαίνεται στο σχήμα 5.15. Ο Haggett, βασιζόμενος στο ονομαζόμενο στη στατιστική «Λατινικό τετράγωνο» (*Latin square*), καθόρισε τον αριθμό των δειγμάτων σε επτά (1-7) ως αποτέλεσμα της ανάγκης να γίνονται μετρήσεις των φόρτων για καθεμιά από τις μέρες της εβδομάδας. Ο χρόνος διαιρέθηκε επίσης σε επτά ίσες χρονικές περιόδους (Δευτέρα-Κυριακή) που κατανέμονται ομοιόμορφα σε

όλη τη διάρκεια της ημέρας (α-η). Η επίσκεψη των επτά σταθμών μέτρησης σε διαφορετικές ώρες αλλά με προκαθορισμένη σειρά κάθε μέρα (Πίνακας 5.1) οδηγεί στη λήψη ενός αμερόληπτου δείγματος της χωρο-χρονικής μεταβολής των κυκλοφοριακών φόρτων. Η εφαρμογή αυτή του λατινικού τετραγώνου έχει ως αποτέλεσμα ο μέσος όρος κάθε μέρα να αποτελείται από μια ισορροπημένη αναλογία θέσεων και ωρών, ο μέσος όρος κάθε ώρας να δημιουργείται από μια ισορροπημένη αναλογία θέσεων και ημερών, ενώ ο μέσος όρος των θέσεων να είναι μια ισορροπημένη αναλογία ωρών και ημερών. Το βασικό πλεονέκτημα αυτής της δειγματοληψίας είναι ότι ο χρόνος για τη δειγματοληψία μειώνεται στο ένα έβδομο (1/7) χωρίς βέβαια να υπολογίζεται ο χρόνος μετάβασης από το ένα σημείο στο άλλο.

	1	2	3	4	5	6	7
Δευτέρα	α	β	γ	δ	ε	ζ	η
Τρίτη	β	ε	α	η	ζ	δ	γ
Τετάρτη	γ	ζ	η	β	δ	α	ε
Πέμπτη	δ	η	ε	ζ	β	γ	α
Παρασκευή	ε	δ	β	γ	α	η	ζ
Σάββατο	ζ	γ	δ	α	η	ε	β
Κυριακή	η	α	ζ	ε	γ	β	δ

Πίνακας 5.1: Καθορισμός «Λατινικού Τετραγώνου»
Χωροχρονικής Δειγματοληψίας



Σχήμα 5.15: Χωρο-χρονική Δειγματοληψία
Πηγή Haggett. *P Locational Analysis* 1975

5.4. ΕΚΤΙΜΗΣΗ ΤΟΥ ΜΕΓΕΘΟΥΣ ΤΟΥ ΔΕΙΓΜΑΤΟΣ

Είναι λοιπόν φανερό από τα παραπάνω ότι η δειγματοληψία ουσιαστικά αφορά έναν συμβιβασμό μεταξύ ακρίβειας και εξοικονόμησης διαθέσιμων (*κόστος, χρόνος κλπ*). Μια παράμετρος μπορεί να υπολογιστεί με 100% ακρίβεια, αν μετρηθεί ή απαριθμηθεί ολόκληρος ο πληθυσμός. Καθώς, όμως, το δείγμα ελαττώνεται, το περιθώριο λάθους στις εκτιμήσεις που γίνονται με βάση αυτό, μεγαλώνει. Γενικά, αν το περιθώριο λάθους που μπορεί να γίνει αποδεκτό σε μια συγκεκριμένη περίπτωση είναι γνωστό, τότε είναι δυνατόν να υπολογιστεί το ελάχιστο μέγεθος του δείγματος που απαιτείται. Με τον τρόπο αυτό εξοικονομείται ο χρόνος που απαιτείται για τη συλλογή ενός μεγάλου σε μέγεθος δείγματος.

Υπάρχουν αρκετοί τρόποι υπολογισμού του ελάχιστου μεγέθους ενός δείγματος. Εδώ θα αναφερθεί ένας μόνο τρόπος προσαρμοσμένος τόσο στην περίπτωση που η δειγματοληψία αναφέρεται σε μετρήσεις όσο και στην περίπτωση που η δειγματοληψία αφορά απαριθμήσεις. Η μέθοδος αυτή απαιτεί τη συλλογή ενός δοκιμαστικού δείγματος που καλό είναι να περιλαμβάνει περίπου 30 παρατηρήσεις. Η δοκιμαστική αυτή δειγματοληψία πρέπει να ακολουθεί τη διαδικασία που περιγράφηκε νωρίτερα και η οποία θα πρέπει να εφαρμόζεται και για το οριστικό δείγμα. Το δοκιμαστικό δείγμα μπορεί, και συνήθως αποτελεί, τμήμα του οριστικού δείγματος, με την προϋπόθεση ότι η επιλογή του δεν έχει επιπτώσεις στην επιλογή των υπόλοιπων παρατηρήσεων για το οριστικό δείγμα.

5.4.1. Μέγεθος Δείγματος για Μετρήσεις

Με βάση τα όσα αναφέρονται στο παράρτημα 5.2 για την εκτίμηση διαστημάτων εμπιστοσύνης, αλλά και τις λεπτομερείς και εμπεριστατωμένες αναφορές πάνω στο θέμα αυτό όλων των βιβλίων στατιστικής, είναι φανερό ότι το περιθώριο λάθους θ στην εκτίμηση της παραμέτρου μ , (*εξίσωση Π.31*) είναι $\pm Z_{\alpha/2}(s/\sqrt{n})$.

Αν η απόλυτη τιμή αυτού του περιθωρίου θεωρηθεί γνωστή, τότε η σχέση γίνεται:

$$n = (Z_{\alpha/2} \cdot s / \theta)^2 \quad (5.1)$$

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

Επομένως, αν για παράδειγμα το δοκιμαστικό δείγμα έχει τυπική απόκλιση ίση με 20 και το περιθώριο λάθους που μπορεί να γίνει αποδεκτό είναι 2 (*στην περίπτωση του μέσου μεγέθους των αγροτεμαχίων αυτό σημαίνει λάθος ± 2 στρέμματα*), τότε για

επίπεδο εμπιστοσύνης 95% το απαιτούμενο μέγεθος του δείγματος θα είναι: $n = (1,96 \cdot 20/2)^2 = 384$. Επομένως, αν το δοκιμαστικό δείγμα είχε 30 παρατηρήσεις, θα απαιτηθούν 354 επιπλέον παρατηρήσεις ώστε να υπάρχει μια εκτίμηση για το επίπεδο εμπιστοσύνης και για το περιθώριο λάθους που να είναι αποδεκτά.

5.4.2. Μέγεθος Δείγματος για Απαριθμήσεις (Αναλογίες)

Στην περίπτωση που οι ανάγκες της μελέτης αναφέρονται σε αναλογίες, με το ίδιο σκεπτικό όπως και στην προηγούμενη περίπτωση, το ελάχιστο μέγεθος ενός δείγματος που μπορεί να χρησιμοποιηθεί για την εκτίμηση της αναλογίας ενός πληθυσμού που ανήκει σε κάποια κατηγορία, δίνεται από τη σχέση:

$$n = 100p \cdot 100q \cdot (Z_{\alpha/2}/\theta)^2 \quad (5.2)$$

όπου: n , $Z_{\alpha/2}$ και θ = τα ίδια όπως και στην σχέση 5.1.

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

Επομένως, αν σε ένα δοκιμαστικό δείγμα 30 κτηρίων τα 15 είναι σε βιομηχανική χρήση ($p=0,5$) και αν δεχτούμε ένα περιθώριο λάθους 2%, τότε για επίπεδο εμπιστοσύνης 95% το απαιτούμενο μέγεθος του δείγματος θα είναι: $50 \times 50(1,96/2)^2 = 2.450$. Δηλαδή θα πρέπει να συνεχιστεί η δειγματοληψία με 2.420 επιπλέον παρατηρήσεις.

5.5. ΑΝΑΦΟΡΕΣ

- Berry, B., and A., Baker, «Geographic Sampling». In B., Berry and D., Marble, **Spatial Analysis**. New Jersey: Prentice-Hall, 1968.
- Cochran, G.W., **Sampling Techniques**. 2nd edition. New York: John Wiley, 1966.
- Cottam, G., Curtis, J., and B., Hale, «Some Sampling Characteristics of a Population of Randomly Dispersed Individuals. **Ecology**, Vol. 34, pp. 741-757, 1953.
- Dixon, C., and B., Leach, **Sampling Methods for Geographic Research**. Norwich: University of East, 1980.
- Hagget, P. «Regional and Local Components in the Distribution of Forested Areas in Southeast Brazil: A Multivariate Approach» **Geographical Journal** vol 130, pp 365-377, 1964.
- Haggett, P., **Locational Analysis in Human Geography**. London: Edward Arnold, 1975.
- Hansen, M.M., Hurwitz, W.N., and W.G., Madow, **Sample Survey Methods and Theory**. New York: John Wiley, 1953.
- Holmes, J., «Problems in Location Sampling». **Annals of the Association of American Geographers**, Vol. 57, pp. 757-758, 1967.
- Kershaw, K., **Quantitative and Dynamic Plant Ecology**. London: Edward Arnold, 1973.
- Krumbein, W.C., «Experimental Design in Earth Sciences». **Transactions of the American Geophysical Union**, Vol. 36, No. 1, 1954.
- Moser, C., and G., Kalton, **Survey Methods in Social Investigation**. London: Heinemann, pp. 61-210, 1971.
- Rowley, G., and P., Shepherd, «A Source of Elements Spatial Data for Town Centre Research in Britain». **Area**, Vol. 8, No. 3, pp. 201-208, 1976.
- Sampford, M., **An introduction to Sampling Theory with Applications to Agriculture**. London: Oliver, 1962.
- Shimwell, D., **The Description and Classification of Vegetation**. London: Sedgwick and Jackson, 1971.
- Stephan, F., and P., McCarthy, **Sampling Opinions**. New York: John Wiley, 1963.
- Stuart, A., **Basic Ideas of Scientific Sampling**. London: Griffin, 1962.
- Wood, W.F., «The Use of Stratified Random Samples in Land Use Study». **Annals of the Association of American Geographers**, Vol. 45, pp. 350, 1955.
- Yates, F., **Sampling Methods for Censuses and Surveys**. London: Griffin, 1953.

ΠΑΡΑΡΤΗΜΑ 5.1: ΠΙΝΑΚΑΣ ΤΥΧΑΙΩΝ ΑΡΙΘΜΩΝ

08210	32973	08003	54512	64863	78634	36344	73293	77660	90199
56822	81546	04735	15228	37475	79057	48749	89898	95554	39734
89182	97110	37211	11688	69132	97318	83419	82573	72506	55631
72674	35938	58483	08612	23129	95675	47142	74290	97894	17531
90431	35286	24567	05594	75288	52097	07687	25548	36274	59709
26899	07988	28526	97601	03588	46466	55143	65558	25847	59878
18345	22317	08035	94062	39380	80045	80231	29109	50908	15304
59231	50655	21753	70621	91045	11318	73098	42457	45905	58889
25446	92080	34212	70395	64493	86192	96633	11311	62422	32794
64825	37691	90465	51579	83918	22217	55262	47065	36168	65350
56902	23614	90057	12880	89522	81453	41294	14603	39279	54730
94781	64586	97431	01310	04978	25833	61253	48959	53841	12864
55140	32776	03346	61125	91434	84200	73664	45914	90551	83073
71284	19521	04812	76028	20044	78244	58015	05867	53847	04736
12938	37575	74532	60769	26928	29666	84318	21283	17537	31834
74888	92697	43016	56171	27506	72769	60939	50508	37332	77234
92592	82165	90227	88560	86908	32283	83958	92590	97910	40983
13741	37073	04809	23138	34852	45986	73823	08534	45867	18800
84984	54147	59332	70137	24535	17957	21012	51279	67206	63087
60528	16871	20195	14420	27684	35554	90414	80832	86089	61328
74340	27455	43144	85543	85639	91044	58951	85251	69761	57118
22071	81837	14952	50079	08851	56337	76769	83059	23254	73901
46002	37283	12700	60802	93248	12585	88835	97516	69585	78330
08389	69818	88126	63220	09217	89809	41827	63143	94194	69001
80520	76260	83873	30245	69153	36222	71188	50184	69731	00382
86956	16358	66811	24846	57071	40649	92444	63235	89322	99933
53136	21244	18794	69981	67197	20656	01157	43212	71717	36254
36225	08162	77121	36757	48280	01627	42182	90286	33607	39507
94504	39905	37479	59621	90903	25258	02601	40884	18925	19143
62903	81216	82151	05623	29557	17355	75959	26697	92117	50471
49684	94165	26723	59644	97156	95754	14613	62639	19567	24652
74368	66450	66013	32765	29673	05637	18535	59269	40732	90216
65947	65189	25064	99907	41696	83361	05711	44600	31021	34714
48562	64985	78945	43273	93601	41398	47898	87513	71023	77141
56879	44941	45205	84341	92222	50193	34172	63957	65014	93068
56352	13144	30800	60403	84654	28738	48297	30614	71784	23440
02272	83016	25229	43692	36151	96505	73151	82126	30327	15817
33209	67546	92428	91522	75736	59065	00668	43097	75282	81904
04042	27494	05204	86656	04755	03880	91082	33651	63806	19334
06255	03873	64275	81962	00650	39269	73457	07644	21562	67255
02637	88632	65996	41115	33863	95116	66522	55347	30163	50594
66788	56790	14606	63123	05831	14117	04288	63610	48924	01827
75185	54242	23276	86228	68663	31051	10832	72330	70330	91769
97678	41209	34692	03790	01489	45480	18809	16099	52339	41146
98264	76970	35868	18540	25973	28110	52129	77706	18505	89583
85483	81260	62113	64994	60821	40518	94273	64432	48779	05310
11364	56369	72114	94681	74135	49670	21072	78489	39630	89270
52663	97730	76199	52766	34479	03055	92954	43490	39790	46437
86276	71122	87556	29504	32565	48447	08719	83014	23615	14317
92295	34604	61004	19707	97387	90740	48389	57337	76910	82215

Πηγή: «A Million Random Digits with 100,000 Normal Deviates» by the RAND Corp.

ΠΑΡΑΡΤΗΜΑ 5.2: ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΣΤΑΤΙΣΤΙΚΗΣ

Π.2.1. ΟΡΙΣΜΟΙ ΠΙΘΑΝΟΤΗΤΩΝ

Π.2.1.1. Πεδίο Δείγματος

Το σύνολο όλων των πιθανών και ανεξάρτητων αποτελεσμάτων ενός απλού πειράματος ονομάζεται πεδίο δείγματος.

Π.2.1.2. Συνάρτηση Πιθανότητας

Δεδομένου ενός πεδίου δείγματος S και ενός συνόλου γεγονότων A στο πεδίο αυτό, μια συνάρτηση πιθανότητας σχετίζει κάθε γεγονός A με έναν πραγματικό αριθμό $P(A)$, την πιθανότητα του γεγονότος A , έτσι ώστε να ισχύουν τα παρακάτω αξιώματα:

$$P(A) \geq 0 \quad \forall A \quad (\text{Π.1})$$

$$P(S) = 1 \quad (\text{Π.2})$$

Αν υπάρχει ένα μετρήσιμο σύνολο γεγονότων $(A_1, A_2, \dots, A_n)$ και αν αυτά τα γεγονότα είναι ξένα μεταξύ τους, τότε:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n) \quad (\text{Π.3})$$

Π.2.1.3. Τυχαία Μεταβλητή

Αν X αντιπροσωπεύει μια συνάρτηση που σχετίζει έναν πραγματικό αριθμό με καθένα από τα στοιχειώδη γεγονότα που ορίζεται στο πεδίο δείγματος S , τότε η X ονομάζεται τυχαία μεταβλητή. Εναλλακτικά: μια ποσότητα που συμβολίζεται με X , ονομάζεται τυχαία μεταβλητή αν για κάθε πραγματικό αριθμό a υπάρχει μια πιθανότητα $P(X \leq a)$, δηλαδή η X παίρνει μια τιμή που είναι μικρότερη ή ίση με a .

Π.2.1.4. Συνάρτηση Κατανομής Τυχαίας Μεταβλητής ή Κατανομή Πιθανοτήτων

Κάθε δήλωση μιας συνάρτησης πιθανότητας που έχει ένα σύνολο ξένων μεταξύ τους γεγονότων ως πεδίο ορισμού, καλείται συνάρτηση κατανομής. Εναλλακτικά: μια συνάρτηση κατανομής, για τα στοιχειώδη γεγονότα στο πεδίο δείγματος S , καθορίζει ένα σύνολο πιθανοτήτων ή μια κατανομή πιθανοτήτων για κάθε τυχαία μεταβλητή ορισμένη στο S .

Π.2.1.5. Διακριτή Μεταβλητή

Αν μια τυχαία μεταβλητή μπορεί να πάρει μόνο ένα πεπερασμένο και μετρήσιμο σύνολο από τιμές λέγεται διακριτή.

Π.2.1.6. Συνεχής Μεταβλητή

Μια τυχαία μεταβλητή ονομάζεται συνεχής αν για κάθε ζεύγος τιμών u και v :

$$\text{αν } P(X \leq u) < P(X \leq v), \text{ ισχύει } P(u < X < v) > 0 \quad (\text{Π.4})$$

Π.2.1.7. Αναμενόμενη Τιμή Τυχαίας Μεταβλητής

Αν η κατανομή της τυχαίας μεταβλητής είναι διακριτή, η αναμενόμενη τιμή ορίζεται:

$$E(X) = \sum_x x P(X = x) = \sum_x x P(x) \quad (\text{Π.5})$$

$$\text{όπου: } \sum_x P(x) = 1$$

Αν η κατανομή είναι συνεχής, η αναμενόμενη τιμή ορίζεται:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad (\text{Π.6})$$

$$\text{όπου: } \int_{-\infty}^{\infty} f(x) dx = 1$$

Π.2.1.8. Δείκτες Κεντρικότητας

Π.2.1.8.1. Μέσος Όρος: Ο μέσος όρος δείχνει την τιμή γύρω από την οποία οι αποκλίσεις προς τη μια κατεύθυνση είναι ακριβώς ίσες με αυτές προς την άλλη κατεύθυνση. Δηλαδή, η τάση των τιμών μιας κατανομής που διαφέρουν από τη μέση τιμή σε μια κατεύθυνση, εξισορροπείται ακριβώς με την τάση να διαφέρουν από την άλλη πλευρά. Ακόμη:

$$\mu = E(X).$$

Π.2.1.8.2. Διάμεσος: Ως διάμεσος ορίζεται η τιμή της Τυχαίας Μεταβλητής X , που είναι στο μέσον (σε σχέση με τις πιθανότητες) μεταξύ της μικρότερης και της μεγαλύτερης δυνατής τιμής $[P(X \leq a) \geq 0,5 \text{ και } P(X \geq a) \geq 0,5]$.

Π.2.1.8.3. Κορυφή: Ορίζεται η τιμή όπου η πιθανότητα της X , $P(X)$, παίρνει τη μεγαλύτερη τιμή.

Π.2.1.9. Δείκτες Διασποράς

Π.2.1.9.1. Διασπορά (σ^2): Είναι ίση με την αναμενόμενη τιμή της τετραγωνισμένης διαφοράς από τη μέση τιμή:

$$\sigma^2 = E(X - \mu)^2 \quad (\text{Π.7})$$

Π.2.1.9.2. Τυπική Απόκλιση: Είναι η τετραγωνική ρίζα της διασποράς και είναι ένας δείκτης διασποράς σε μονάδες της αρχικής μέτρησης

$$\sigma = \sqrt{\sigma^2} = \sqrt{E(X - \mu)^2} \quad (\text{Π.8})$$

Π.2.2. ΟΡΙΣΜΟΙ ΣΤΑΤΙΣΤΙΚΗΣ

Π.2.2.1. Δείγμα

Δεδομένου κάποιου πειράματος και ενός πεδίου δείγματος από στοιχειώδη γεγονότα, το σύνολο των αποτελεσμάτων N ξεχωριστών επιλογών είναι το δείγμα.

Π.2.2.2. Τυχαία δειγματοληψία

Μια μέθοδος δημιουργίας δειγμάτων, έτσι ώστε κάθε ξεχωριστό δείγμα του ίδιου μεγέθους N έχει ακριβώς την **ίδια πιθανότητα** επιλογής, ονομάζεται απλή τυχαία δειγματοληψία.

Π.2.2.3. Ανεξάρτητη τυχαία δειγματοληψία

Στοιχειώδη γεγονότα δειγματίζονται τυχαία, όταν η επιλογή ενός τέτοιου γεγονότος δεν έχει απολύτως **καμία σχέση** με την πιθανότητα επιλογής του ίδιου ή άλλου στοιχειώδους γεγονότος σε μια μεταγενέστερη επιλογή.

Π.2.2.4. Πληθυσμός

Πληθυσμός σημαίνει το σύνολο των δυνατών μονάδων για παρατήρηση.

Π.2.2.5. Παράμετρος

Πληθυσμιακοί δείκτες όπως μ και σ^2 ονομάζονται **παράμετροι**. Πιο συγκεκριμένα, παράμετρος είναι μια τιμή που εισέρχεται ως αυθαίρετη σταθερή στο συγκεκριμένο συναρτησιακό κανόνα μιας συνάρτησης κατανομής.

Π.2.2.6. Κατανομή Συχνοτήτων

Κάθε απεικόνιση της σχέσης ανάμεσα σε ένα σύνολο από ξένες μεταξύ τους ομάδες μετρήσεων και της συχνότητάς τους ονομάζεται κατανομή συχνοτήτων.

Π.2.2.7. Μέσος Όρος Δείγματος

Δίνεται από τους τύπους:

$$M = \frac{\sum_{i=1}^N X_i}{N} \quad (\text{Π.9})$$

ή στην περίπτωση κατανομής συχνοτήτων ομάδων:

$$M = \frac{\sum xf}{N} \quad (\text{Π.10})$$

όπου: X_i = παρατηρούμενη μεταβλητή
 N = μέγεθος δείγματος
 x = μέση τιμή του διαστήματος της ομάδας
 f = συχνότητα της ομάδας.

Π.2.2.8. Διασπορά Δείγματος

Δίνεται από τους τύπους:

$$s^2 = \frac{\sum_{i=1}^N (X_i - M)^2}{N} \quad (\text{Π.11})$$

ή

$$s^2 = \frac{N \sum X_i^2 - (\sum X_i)^2}{N^2} \quad (\text{Π.12})$$

Π.2.2.9. Δειγματοληπτικές κατανομές

Η δειγματοληπτική κατανομή είναι μια θεωρητική κατανομή που δείχνει τη συναρτησιακή σχέση ανάμεσα σε όλες τις πιθανές τιμές ενός **δείκτη** από N γεγονότα που επιλέγονται τυχαία και της πιθανότητας που σχετίζεται με κάθε τιμή του δείκτη αυτού για το σύνολο των πιθανών δειγμάτων μεγέθους N ενός συγκεκριμένου πληθυσμού.

Π.2.2.10. Δείκτες Δειγματοληπτικής Κατανομής

Π.2.2.10.1. Μέσος Όρος: Δίνεται από τους τύπους:

$$\mu_G = E(G) = \sum gP(g) \quad \text{για διακριτές τυχαίες μεταβλητές} \quad (\text{Π.13})$$

$$\mu_G = E(G) = \int_{-\infty}^{\infty} gf(g)dg \quad \text{για συνεχείς τυχαίες μεταβλητές} \quad (\text{Π.14})$$

Π.2.2.10.2. Διασπορά: Δίνεται από τον τύπο

$$\sigma^2 = E(G - \mu_G)^2 = E(G^2) - [E(G)]^2 \quad (\text{Π.15})$$

Π.2.2.11. Οι Τρεις Βασικές Κατανομές

Π.2.2.11.1. Συνάρτηση Πιθανοτήτων (Πληθυσμός): Περιγράφει τη σχέση ανάμεσα σε μια τυχαία μεταβλητή και τις πιθανότητες της.

Π.2.2.11.2. Κατανομή Συχνοτήτων (Δείγμα): Περιγράφει ομάδες παρατηρήσεων με τις συχνότητες τους σε ένα τυχαίο δείγμα.

Π.2.2.11.3. Δειγματοληπτική Κατανομή (Δείκτες Δειγμάτων): Περιγράφει τη σχέση ανάμεσα σε ένα δείκτη δείγματος ορισμένου μεγέθους και τις πιθανότητες του.

Π.2.3. ΟΡΙΣΜΟΙ ΕΚΤΙΜΗΣΕΩΝ

Π.2.3.1. Σημειακές εκτιμήσεις

Κάθε εκτιμητής σημείου πρέπει να είναι:

- Αμερόληπτος εκτιμητής (ο M είναι αμερόληπτος εκτιμητής του μ , αλλά όχι ο s^2 του σ^2)
- Συμβατός εκτιμητής
- Σχετικά αποτελεσματικός εκτιμητής
- Επαρκής εκτιμητής.

Π.2.3.2. Μέθοδοι Εκτίμησης Σημειακών Εκτιμητών

Οι βασικές μέθοδοι εύρεσης εκτιμητών σημείων είναι οι εξής:

- Μέθοδος της Μέγιστης Πιθανοφάνειας
- Μέθοδος των Ροπών
- Bayesian Μέθοδος
- Μέθοδος των Ελάχιστων Τετραγώνων.

Π.2.3.3. Διάστημα Εμπιστοσύνης

Διάστημα εμπιστοσύνης είναι μια περιοχή τιμών με δεδομένη την πιθανότητα ότι καλύπτει την πραγματική για τον πληθυσμό τιμή.

Π.2.3.3.1. Η Κανονική Κατανομή, Δειγματοληπτική Κατανομή Μέσων Όρων Δείγματος: Αν ένας πληθυσμός έχει κανονική κατανομή, τότε και ο δειγματικός μέσος όρος M έχει κανονική κατανομή (θεώρημα του κεντρικού ορίου). Επομένως η

κανονική κατανομή και μάλιστα στην τυποποιημένη της (*standardized*) μορφή, αποτελεί τη δειγματοληπτική κατανομή για την εύρεση του διαστήματος εμπιστοσύνης εκτίμησης του μ από τον M .

Π.2.3.3.2. Η Κατανομή t , Δειγματοληπτική Κατανομή Μέσων Όρων Δείγματος όταν σ^2 Αγνώστη: Αν η σ^2 δεν είναι γνωστή τότε μπορεί να χρησιμοποιηθεί ένας εκτιμητής της, που είναι η δειγματική s^2 .

Επειδή είναι γνωστό ότι:

$$s^2 = \frac{\sum (X_i - M)^2}{N} \quad (\text{Π. 16})$$

και

$$Z_M = \frac{M - \mu}{\sigma} \quad (\text{Π.17})$$

τότε δημιουργείται μια νέα τυποποιημένη τυχαία μεταβλητή ως εξής:

$$t = \frac{M - \mu}{s/\sqrt{N-1}} \quad \text{ή καλύτερα} \quad (\text{Π.18})$$

$$t = \frac{M - \mu}{s\sqrt{N}} \quad (\sqrt{N} = \text{o αμερόληπτος εκτιμητής της } \sigma^2) \quad (\text{Π. 19})$$

Η κατανομή t επομένως είναι η δειγματοληπτική κατανομή των μέσων όρων δειγμάτων πληθυσμού, του οποίου δεν είναι γνωστή η διασπορά και μπορεί να χρησιμοποιηθεί για την εύρεση του διαστήματος εμπιστοσύνης εκτίμησης του μ από το M όταν η σ είναι άγνωστη.

Η Z_M διαφέρει από την t γιατί στην πρώτη ο αριθμητής είναι τυχαία μεταβλητή και ο παρανομαστής σταθερός, ενώ στη δεύτερη αριθμητής και παρονομαστής είναι τυχαίες μεταβλητές.

Π.2.3.3.3. Η Διωνυμική Κατανομή, Δειγματοληπτική Κατανομή Διχοτομικών Μεταβλητών: Αν η πιθανότητα επιλογής ενός στοιχειώδους γεγονότος που έχει ένα χαρακτηριστικό είναι p και η πιθανότητα να μην έχει αυτό το χαρακτηριστικό ορίζεται ως q (έτσι ώστε $p+q = 1$), τότε σε ένα τυχαίο δείγμα μεγέθους n η συχνότητα αυτού του χαρακτηριστικού είναι np . Έχει αποδειχθεί ότι η δειγματοληπτική κατανομή αυτών των συχνοτήτων είναι η διωνυμική κατανομή, η οποία προσεγγίζει την κανονική κατανομή με μέσο όρο np και τυπικό σφάλμα $\sqrt{npq}$. Επομένως, η διωνυμική κατανομή μπορεί να χρησιμοποιηθεί για την εύρεση του διαστήματος εμπιστοσύνης εκτίμησης του p από διχοτομικές δειγματικές συχνότητες.

Π.2.3.3.4. Η Κατανομή χ^2 , Δειγματοληπτική Κατανομή της σ^2 . Ο λόγος Ns^2/σ^2 είναι μια χ^2 μεταβλητή με $N-1$ βαθμούς ελευθερίας, η καλύτερα ο λόγος $(N-1)s^2$ προς σ^2 είναι μια τυχαία μεταβλητή που κατανέμεται σαν χ^2 με $N-1$ βαθμούς ελευθερίας. Επομένως:

$$P\left[\chi^2_{(1-\alpha/2, N-1)} \leq \frac{(N-1)\hat{s}^2}{\sigma^2} \leq \chi^2_{(\alpha/2, N-1)}\right] = 1-\alpha \quad (\text{Π.20})$$

ή

$$P\left[\frac{(N-1)\hat{s}^2}{\chi_{(\alpha/2, N-1)}} \leq \sigma^2 \leq \frac{(N-1)\hat{s}^2}{\chi_{(1-\alpha/2, N-1)}}\right] = 1-\alpha \quad (\text{Π.21})$$

Η κατανομή χ^2 επομένως μπορεί να χρησιμοποιηθεί για την εύρεση του διαστήματος εμπιστοσύνης εκτίμησης της σ^2 από την s^2 .

Π.2.3.3.5. Η Κατανομή F , Δειγματοληπτική Κατανομή Λόγου Διασπορών: Μια τυχαία μεταβλητή που σχετίζεται με το λόγο δύο ανεξάρτητων χ^2 μεταβλητών, διαιρούμενης καθεμιάς με τους βαθμούς ελευθερίας της, ονομάζεται λόγος F και ακολουθεί μια κατανομή F . Επειδή:

$$\frac{(N-1)s^2}{\sigma^2} = \chi^2_{(N-1)} \quad \text{ή} \quad \frac{\hat{s}_1^2}{\sigma_1^2} = \frac{\chi_{(v_1)}^2}{v_1} \quad \text{και ανάλογα} \quad (\text{Π.22})$$

$$\frac{s_2}{\sigma_2} = \frac{\chi_{(v_2)}}{v_2} \quad \text{τότε} \quad F = \frac{\hat{s}_1^2}{\hat{s}_2^2} = \frac{[\chi_{v_1}^2 / v_1]}{[\chi_{v_2}^2 / v_2]} = \frac{\hat{s}_1^2 / \sigma_1^2}{\hat{s}_2^2 / \sigma_2^2} \quad \text{και άρα:} \quad (\text{Π.23})$$

$$P\left[\frac{1}{F_{(\alpha/2, N_2-1, N_1-1)}} \leq \frac{\hat{s}_1^2 / \sigma_1^2}{\hat{s}_2^2 / \sigma_2^2} \leq F_{(\alpha/2, N_2-1, N_1-1)}\right] = 1-\alpha \quad \text{ή} \quad P\left[\frac{1}{F_{(\alpha/2, N_2-1, N_1-1)}} \cdot \frac{\hat{s}_2^2}{\hat{s}_1^2} \leq \frac{\sigma_2}{\sigma_1} \leq F_{(\alpha/2, N_2-1, N_1-1)} \frac{\hat{s}_2^2}{\hat{s}_1^2}\right] = 1-\alpha \quad (\text{Π.24})$$

Η κατανομή F επομένως μπορεί να χρησιμοποιηθεί για την εύρεση του διαστήματος εμπιστοσύνης εκτίμησης λόγου διασπορών.

Π.2.4. ΕΚΤΙΜΗΣΕΙΣ ΔΙΑΣΤΗΜΑΤΩΝ ΕΜΠΙΣΤΟΣΥΝΗΣ ΓΙΑ ΤΟ ΜΕΓΕΘΟΣ ΔΕΙΓΜΑΤΟΣ

Όπως αναφέρθηκε προηγουμένως, τα δείγματα χρησιμοποιούνται για να επιτυγχάνονται εκτιμήσεις χαρακτηριστικών του πληθυσμού, όταν δεν είναι εύκολο ή πρακτικό να μετρηθούν ή να απαριθμηθούν σε ολόκληρο τον πληθυσμό. Διεθνώς έχει επικρατήσει να χρησιμοποιείται ο όρος **παράμετρος πληθυσμού** (*population parameter*) για να δηλωθεί το χαρακτηριστικό ενός πληθυσμού και **εκτιμητής δείγματος** (*sample statistic*) για να δηλωθεί το χαρακτηριστικό ενός δείγματος. Γενικά, ο εκτιμητής ενός δείγματος (π.χ. το μέσο μέγεθος 100 τυχαία δειγματοληφθέντων αγροτεμαχίων στο Νομό Μαγνησίας) δίνει τη μόνη υπάρχουσα εκτίμηση για την αντίστοιχη παράμετρο του πληθυσμού (π.χ. το μέσο μέγεθος των αγροτεμαχίων στο Νομό Μαγνησίας). Οποιοδήποτε δείγμα, ανεξάρτητα πόσο αμερόληπτο είναι, να δώσει τιμές για τη μέση τιμή (M) και την τυπική απόκλιση (s), που να είναι ακριβείς εκτιμητές των αντιστοίχων παραμέτρων του πληθυσμού (μ , σ). Είναι το ίδιο απίθανο δύο δείγματα, που λαμβάνονται από τον ίδιο πληθυσμό (*δειγματοληπτικό υπόβαθρο*), να έχουν τις ίδιες τιμές για το M και s . Αντίθετα, αν ληφθεί ένας μεγάλος αριθμός δειγμάτων από τον ίδιο πληθυσμό, το πιο πιθανό είναι ότι κάθε δείγμα θα δημιουργήσει μια διαφορετική μέση τιμή και μια διαφορετική τυπική απόκλιση. Η κατανομή όλων αυτών των εκτιμητών των δειγμάτων (π.χ. η μέση τιμή των δειγμάτων που λαμβάνονται από ένα πληθυσμό) είναι η **δειγματοληπτική κατανομή** τους. Αν τα δείγματα είναι αμερόληπτα, είναι λογικό να αναμένεται ότι η μέση τιμή τους, όπως και κάθε άλλος εκτιμητής των δειγμάτων, θα κατανέμεται συμμετρικά γύρω από την αντίστοιχη παράμετρο του πληθυσμού και επομένως θα ήταν δυνατό να διαμορφωθεί μια κάποια αντίληψη για την πραγματική τιμή τους.

Βέβαια στην πράξη δεν λαμβάνονται πολλά δείγματα αλλά μόνο ένα και οι παράμετροι του πληθυσμού που δεν είναι γνωστοί, εκτιμώνται από το μοναδικό αυτό δείγμα. Η πεμπτουσία επομένως της εκτίμησης στην στατιστική είναι ακριβώς η **διαδικασία από την κατανομή συχνοτήτων (κατανομή του δείγματος) στη συνάρτηση κατανομής του πληθυσμού με τη βοήθεια μιας δειγματοληπτικής κατανομής**.

Π.2.4.1. Εκτιμήσεις Διαστημάτων Εμπιστοσύνης για Μετρήσεις

Η διαδικασία αυτή εκτίμησης, τουλάχιστον για τις περιπτώσεις που αντιμετωπίζουν οι περισσότεροι ερευνητές, βασίζεται σε ένα θεμελιώδες θεώρημα που ονομάζεται το **θεώρημα του κεντρικού ορίου** (*central limit theorem*). Το θεώρημα αυτό λέει ότι αν θεωρηθεί ότι λαμβάνονται όλα τα δυνατά δείγματα του

ίδιου μεγέθους από ένα συγκεκριμένο πληθυσμό, η δειγματοληπτική κατανομή των μέσων τιμών θα προσεγγίζει την κανονική κατανομή, που η μέση τιμή της είναι η μέση τιμή του πληθυσμού και το τυπικό σφάλμα (έτσι ονομάζεται για λόγους διαφοροποίησης η τυπική απόκλιση των δειγματοληπτικών κατανομών) θα δίνεται από τη σχέση:

$$\sigma_M = \sigma / \sqrt{n} \quad (\text{Π.25})$$

όπου: σ_M = το τυπικό σφάλμα της δειγματοληπτικής κατανομής των $\mathbf{M}$

σ = η τυπική απόκλιση του πληθυσμού

n = το μέγεθος του δείγματος.

Αν ληφθεί η δειγματοληπτική αυτή κατανομή στην τυποποιημένη της μορφή τότε:

$$Z_M = \frac{M - \mu}{\sigma_M} = \frac{M - \mu}{\sigma / \sqrt{n}} \quad (\text{Π.26})$$

Είναι γνωστό ότι η τυποποιημένη τυχαία μεταβλητή για επίπεδο εμπιστοσύνης **(100 α) %** παίρνει τιμές μεταξύ $Z_{\alpha/2}$ και $-Z_{\alpha/2}$ (από τους πίνακες της τυποποιημένης κανονικής κατανομής). Δηλαδή:

$$P(-Z_{\alpha/2} \leq Z_M \leq Z_{\alpha/2}) = 100 - \alpha \quad (\text{Π.27})$$

Αντικαθιστώντας τον Z_M από την εξίσωση Π.26, τότε:

$$P\left(-Z_{\alpha/2} \leq \frac{M - \mu}{\sigma_M} \leq Z_{\alpha/2}\right) = 100 - \alpha \quad \text{ή} \quad (\text{Π.28})$$

$$P(M - Z_{\alpha/2} \sigma_M \leq \mu \leq M + Z_{\alpha/2} \sigma_M) = 100 - \alpha \quad (\text{Π.29})$$

Επομένως, όταν γίνεται εκτίμηση της μέσης τιμής του πληθυσμού (μ) με βάση τη μέση τιμή του δείγματος $\mathbf{M}$, τότε κάποιος μπορεί να είναι **(100-α)%** σίγουρος ότι η πραγματική τιμή του πληθυσμού θα βρίσκεται στο διάστημα εμπιστοσύνης:

$$M \pm Z_{\alpha/2} \sigma_M = M \pm Z_{\alpha/2} (\sigma / \sqrt{n}) \quad (\text{Π.30})$$

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

Η εύρεση του διαστήματος εμπιστοσύνης από τη σχέση Π.26 προϋποθέτει ότι είναι γνωστή η τυπική απόκλιση σ του πληθυσμού. Επειδή συνήθως η σ δεν είναι γνωστή, για μεγάλα δείγματα (μεγαλύτερο των 30 παρατηρήσεων) χρησιμοποιείται ένας εκτιμητής της σ , που είναι η τυπική απόκλιση του δείγματος $\mathbf{s}$, και το **(100-α)%** διάστημα εμπιστοσύνης δίνεται από τη σχέση:

$$M \pm Z_{\alpha/2} (s / \sqrt{n}) \quad (\text{Π.31})$$

Στην περίπτωση που το δείγμα είναι μικρό (μικρότερο των 30) και με αποδοχή της υπόθεσης ότι η κατανομή του πληθυσμού είναι κανονική, ως εκτιμητής του σ δεν

χρησιμοποιείται το s αλλά η έκφραση: $s\sqrt{n/(n-1)}$. Το $(100-\alpha)\%$ διάστημα εμπιστοσύνης, στην περίπτωση αυτή, δίνεται από τη σχέση:

$$M \pm Z_{\alpha/2} \left(s / \sqrt{n-1} \right) \quad (\text{Π.32})$$

Συνήθως, όμως, στην περίπτωση που το δείγμα είναι μικρό, αντί για την κανονική κατανομή, χρησιμοποιείται η t κατανομή και το διάστημα εμπιστοσύνης δίνεται από τη σχέση:

$$M \pm t_{\alpha/2} \left(s / \sqrt{n} \right) \quad (\text{Π.33})$$

όπου: $t_{\alpha/2}$ = η τιμή από τους πίνακες της t κατανομής που αντιστοιχεί στην αθροιστική πιθανότητα $100-(\alpha/2)$.

Για να γίνουν πιο κατανοητά όσα περιληπτικά διατυπώθηκαν παραπάνω, θα ήταν σκόπιμο να εξεταστεί ένα υποθετικό παράδειγμα. Ας υποτεθεί ότι είναι ανάγκη να εκτιμηθεί το μέσο μέγεθος των αγροτεμαχίων του νομού Μαγνησίας. Καταρχάς λαμβάνεται μια τυχαία δειγματοληψία 100 αγροτεμαχίων, που η ανάλυση δείχνει ότι έχουν μέσο μέγεθος (σε στρέμματα) $M = 50$ και μια τυπική απόκλιση $s = 10$. Με βάση τις τιμές αυτές βρίσκουμε (από τη σχέση Π.25) ότι το τυπικό σφάλμα $\sigma_M = 10/\sqrt{100} = 1$ (χρησιμοποιώντας λόγω μεγέθους δείγματος το s ως εκτιμητή του σ). Αν τώρα είναι επιθυμητό να υπάρχει μια σιγουριά 95% για το αποτέλεσμα ($\alpha = 5\%$) τότε η τιμή $Z_{.975}$ (από τους πίνακες της κανονικής κατανομής) βρίσκεται ότι είναι 1,96. Επομένως, από τη σχέση Π.27 βρίσκεται ότι το διάστημα $M \pm Z_{\alpha/2} (s/\sqrt{n})$ είναι ίσο με $50 \pm 1,96$. Δηλαδή, η πιθανότητα το μέσο μέγεθος των αγροτεμαχίων του νομού Μαγνησίας να είναι μεταξύ 48,04 και 51,96 στρέμματα είναι 95%. Όμοια, υπάρχει μια σιγουριά 99% ότι το μέσο μέγεθος των αγροκτημάτων είναι μεταξύ 52,58 και 57,42 στρέμματα ($Z_{.995} = 2,58$).

Π.2.4.2. Εκτιμήσεις Διαστημάτων Εμπιστοσύνης από Απαριθμήσεις

Στα προηγούμενα κεφάλαια εξετάσθηκαν στατιστικές εκτιμήσεις από δείγματα, που οι επιμέρους παρατηρήσεις τους ήταν κάποια μέτρηση (π.χ. το μέσο μέγεθος των αγροτεμαχίων). Συχνά όμως, κυρίως σε θέματα που ασχολούνται οι μηχανικοί, οι παρατηρήσεις είναι απαριθμήσεις και όχι μετρήσεις. Για παράδειγμα, υπάρχει μεγάλο ενδιαφέρον να γίνει γνωστός ο αριθμός των κτηρίων μιας πόλης, που χρησιμοποιούνται από βιομηχανίες. Επομένως, σε ένα δείγμα των κτηρίων αυτής της πόλης, οι παρατηρήσεις θα είναι κτίσματα με βιομηχανική χρήση και με μη βιομηχανική χρήση, που όταν αθροιστούν (π.χ. από ένα δείγμα 200 κτηρίων τα 18 είναι σε βιομηχανική χρήση), τότε και η αναλογία των κτηρίων σε βιομηχανική χρήση

στο δείγμα θα μπορεί να υπολογιστεί (9%). Όμως, με τις τεχνικές που περιληπτικά αναφέρθηκαν προηγουμένως, δεν είναι δυνατόν να γίνουν εκτιμήσεις, γιατί η τυπική απόκλιση για ένα τέτοιο δείγμα απαριθμήσεων δεν μπορεί να υπολογιστεί και επομένως ούτε και το τυπικό σφάλμα και φυσικά ούτε το διάστημα εμπιστοσύνης. Βασικά, απαιτείται μια διαφορετική δειγματοληπτική κατανομή **-όχι διαφορετική μεθοδολογία-** για να γίνουν εκτιμήσεις για τα πληθυσμιακά μεγέθη. Η δειγματοληπτική κατανομή που μπορεί να βοηθήσει σε τέτοιες περιπτώσεις είναι η **δυωνυμική**. Πραγματικά, η δυωνυμική κατανομή σχετίζεται με την επανάληψη τυχαίων γεγονότων που μπορούν να λάβουν δύο μόνο μορφές (π.χ. *αγροτική γη - μη αγροτική γη, κτίσμα ενός ορόφου - πολυώροφο κτίσμα κλπ*).

Επομένως, όσον αφορά την εκτίμηση τέτοιων απαριθμήσεων (π.χ. *ο αριθμός κτισμάτων σε βιομηχανική χρήση*), μπορεί να χρησιμοποιηθεί η σχέση Π.26 αντικαθιστώντας τη μέση τιμή **M** με **np** και το τυπικό σφάλμα **σ_M** με $\sqrt{npq}$. Ως αποτέλεσμα, το διάστημα εμπιστοσύνης θα δίνεται από τη σχέση:

$$np \pm Z_{\alpha/2} \sqrt{npq} \quad (\text{Π.34})$$

Όμως, όταν επιχειρούνται στατιστικές εκτιμήσεις για απαριθμήσεις, συνήθως το ενδιαφέρον εστιάζεται περισσότερο στις αναλογίες παρά στον ακριβή αριθμό. Δηλαδή υπάρχει μεγαλύτερη ανάγκη να γίνει γνωστή η αναλογία των κτισμάτων σε βιομηχανική χρήση σε σύγκριση με το σύνολο των κτισμάτων, παρά ο ακριβής αριθμός αυτών των κτισμάτων. Πρέπει, επομένως, να εκφραστεί η μέση τιμή και το τυπικό σφάλμα της δειγματοληπτικής κατανομής ως ποσοστό του δείγματος. Αυτό επιτυγχάνεται με τον πολλαπλασιασμό των προηγουμένων εκφράσεων με **100/n**.

Έτσι η μέση τιμή **np** γίνεται **np x (100/n) = 100p** τοις εκατό του **n**, και το τυπικό σφάλμα $\sqrt{npq}$ γίνεται $\sqrt{npq} \times (100/n) = \sqrt{100p \times 100q/n}$. Αν τα μεγέθη αυτά αντικαταστήσουν τα αντίστοιχά τους στη σχέση Π.29, το διάστημα εμπιστοσύνης δίνεται πλέον σε επί τοις εκατό αναλογίες ως εξής:

$$np \pm Z_{\alpha/2} \sqrt{100p \cdot 100q/n} \quad (\text{Π.35})$$