

## ΚΕΦΑΛΑΙΟ 1

### ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΠΛΗΘΥΣΜΩΝ

#### 1. Εισαγωγή

Στις σύγχρονες ποσοτικές έρευνες και μελέτες οικολογικής, περιβαλλοντικής και γεωργικής φύσεως, ιδίως σ' εκείνες που πραγματοποιούνται σε συνθήκες υπαίθρου, χρειάζονται καλές μετρήσεις των στοιχείων ή χαρακτηριστικών του συστήματος που μελετάμε.

Ειδικά στις περιπτώσεις ερευνών σε επίπεδο πληθυσμού ή βιοκοινότητας, η απογραφή ή η μέτρηση των σχετικών παραμέτρων όλων των ατόμων του υπό μελέτη –μεγάλου– συνόλου είναι κατά κανόνα αδύνατη ή πολύ δαπανηρή. Περιοριζόμαστε λοιπόν στη λήψη *αντιπροσωπευτικού δείγματος*, το οποίο αν επιλεγεί σωστά (σύμφωνα με τους κανόνες της δειγματοληψίας), μπορεί να χρησιμοποιηθεί για να γίνουν αναφορές στον πληθυσμό ή σ' όποιο σύνολο από το οποίο ελήφθη το δείγμα.

Το θέμα της δειγματοληψίας γενικά είναι αρκετά μεγάλο. Τα όργανα και οι μέθοδοι δειγματοληψίας που χρησιμοποιούνται ποικίλλουν ανάλογα με το είδος των οργανισμών που μελετούμε και τον χώρο (χερσαίος, θαλάσσιος, ποτάμιος, λιμναίος κ.τ.λ.) στον οποίο ζουν. Στο παρόν κεφάλαιο και στα επόμενα τρία θα εξετάσουμε τις βασικές αρχές της δειγματοληψίας και τις πλέον συνήθεις μεθόδους και όργανα που χρησιμοποιούνται. Για εκτενέστερη κάλυψη και μεγαλύτερη εμβάθυνση στο θέμα ο ερευνητής θα πρέπει να ανατρέξει σε εξειδικευμένα εγχειρίδια (Snedecor and Cochran 1967, Seber 1973, Begon 1979, Krebs 1999, Govindarajulu 1999). Σε ορισμένες περιπτώσεις η δειγματοληψία ποσοτικών χαρακτηριστικών των φυτών χρησιμοποιείται για την έμμεση αξιολόγηση (βιοκαταγραφή) των επιπέδων κάποιας περιβαλλοντικής παραμέτρου, όπως π.χ. ατμοσφαιρικής ρύπανσης (Saitanis and Karandinos 2001). Όσον αφορά τις τεχνικές και τα όργανα δειγματοληψίας, ιδιαίτερα σε υδάτινα περιβάλλοντα, το βιβλίο των Southwood and Henderson (2000) είναι αρκετά λεπτομερές. Επίσης χρήσιμο ίσως είναι στον αναγνώστη το Παράρτημα 1 τούτου του βιβλίου.

Πρέπει να τονίσουμε ότι για τον σχεδιασμό μιας καλής δειγματοληψίας με στόχο την εκτίμηση παραμέτρων του πληθυσμού κάποιου είδους (π.χ. πληθυσμιακό μέγεθος, λόγο αρρένων προς θήλεα άτομα, ύψος προσβολής μιας καλλιέργειας

από κάποιο έντομο ή μύκητα, ποσοστό παρασιτισμού κ.τ.λ.), ο ερευνητής θα πρέπει να έχει τις γνώσεις εκείνες της βιολογίας και οικολογίας (σε ποιοτικό έστω επίπεδο) του υπό μελέτη είδους που εμπλέκονται με την τεχνική της δειγματοληψίας την οποία προτίθεται να χρησιμοποιήσει. Για παράδειγμα, οι Duffield et al. (2005) περιγράφουν τη βαθμονόμηση διαφόρων τεχνικών δειγματοληψίας αυγών και προνυμφών του Λεπιδοπτέρου εντόμου *Helicoverpa spp.* σε φυτείες σόγιας. Επίσης, οι Stephens and Losey (2004) συνέκριναν διάφορες μεθόδους δειγματοληψίας των Coccinellidae στο καλαμπόκι. Εάν σε δεδομένη περίπτωση δεν υπάρχει σχετική γνώση, θα πρέπει να αποκτηθεί, με την πραγματοποίηση ενδεχομένως κάποιων παρατηρήσεων και προκαταρκτικών δειγματοληψιών. Ειδικά στην περίπτωση δειγματοληψίας ειδών με μικρούς πληθυσμούς (π.χ. νέα είδη στην περιοχή που πιθανόν να προκαλέσουν επιδημίες), οι Venette et al. (2002) σε ένα άρθρο ανασκόπησης περιγράφουν τις κατάλληλες στατιστικές στρατηγικές της δειγματοληψίας. Βέβαια, κανόνες γενικής εφαρμογής της δειγματοληψίας για όλα τα είδη και πληθυσμούς δεν υπάρχουν. Ενδεικτικά μόνο αναφέρουμε μερικά παραδείγματα.

Έστω ότι η δειγματοληψία γίνεται με ελκυστικά σε παγίδες για την εκτίμηση του πληθυσμού ακμαίων εντόμων κάποιου είδους. Ο ερευνητής θα πρέπει να γνωρίζει εάν προσελκύονται τα άτομα του είδους αυτού ανεξαρτήτως ηλικίας, φύλου, συζευγμένα ή μη κ.τ.λ. Δηλαδή ποιος είναι ο «πληθυσμός» από τον οποίο λαμβάνεται το δείγμα. Εάν η δειγματοληψία γίνεται με μεθόδους που σχετίζονται με την κινητικότητα των ζώων και ιδιαίτερα αν η έρευνα στοχεύει στην καταγραφή των πληθυσμιακών διακυμάνσεων στο χρόνο (π.χ. εποχιακές), ο ερευνητής θα πρέπει να λάβει υπόψη του τις περιβαλλοντικές συνθήκες που ευνοούν ή παρεμποδίζουν την κινητικότητα των ζώων σε κάθε χρονική περίοδο.

Στην εφαρμοσμένη εντομολογία η ζημία της παραγωγής συχνά προκαλείται από τις προνύμφες του εντόμου, η ποσοτική δειγματοληψία των οποίων μπορεί να είναι δύσκολη ή/και χρονοβόρα. Επιπλέον, εάν ο σκοπός της δειγματοληψίας είναι να εκτιμήσουμε την αναμενόμενη ζημία στην καλλιέργειά μας από τις προνύμφες –για να επέμβουμε εάν χρειάζεται–, τότε η δειγματοληψία στο στάδιο της προνύμφης ίσως δε βοηθά. Θα πρέπει να διενεργήσουμε τη δειγματοληψία έγκαιρα στο στάδιο του ακμαίου της προηγούμενης γενιάς (π.χ. με παγίδες φερομόνης), έχοντας όμως προσδιορίσει ήδη ερευνητικά τη συσχέτιση μεταξύ συλλήψεων ακμαίων στις παγίδες και αναμενόμενου πληθυσμιακού μεγέθους των προνυμφών. Τέτοιες έρευνες έχουν πραγματοποιηθεί για διάφορα οικονομικής σημασίας έντομα. Σε μια πρόσφατη (Hillier et al 2004) τέτοια έρευνα μελετήθηκε επί τρία χρόνια στον Καναδά το Λεπιδόπτερο *Grapholita libertina* που προσβάλλει τα *Vaccinium vitis-idaea L.* (κοινώς μυρτιάδα ή μπλούμπερη). Πράγματι, στην περίπτωση αυτή βρέθηκε καλή συσχέτιση μεταξύ του αριθμού

των συλληφθέντων ακμαίων και: (i) του αριθμού των προνυμφών και (ii) του επιπέδου της ζημίας.

Σε μια εμπειριστατωμένη έρευνα 7 ετών στις ΗΠΑ (Raimondo et al. 2004) για το έντομο *Lymantria dispar*, που προσβάλλει διάφορα οπωροφόρα και δασικά δένδρα, μελετήθηκε η συγκριτική αποτελεσματικότητα 4 μεθόδων δειγματοληψίας (συλλήψεις σε φωτοπαγίδες, αριθμός προνυμφών φυλλώματος, αριθμός ωομαζών και συλλήψεις σε “παγίδες καμβά” στον κορμό των δένδρων) και σε 3 κλίμακες χώρου (τεμάχια δειγματοληψίας 200 ha, δασικές συστάδες 200-3.000 ha, και “περιοχή” 200.000 ha). Οι αριθμοί των ακμαίων στις φωτοπαγίδες αξιολογήθηκαν ως καλοί δείκτες του πληθυσμιακού μεγέθους σε σχέση με τις καταμετρήσεις προνυμφών φυλλώματος. Επίσης, αριθμοί προνυμφών φυλλώματος και συλλήψεις στον καμβά ήταν καλοί δείκτες του πληθυσμιακού μεγέθους στις κλίμακες των τεμαχίων και δασικών συστάδων, αλλά μόνο αριθμοί προνυμφών φυλλώματος ήταν καλοί δείκτες και στην κλίμακα της περιοχής.

Οι Μπρούμας και Σουλιώτης (1987) τεκμηρίωσαν σημαντική συσχέτιση μεταξύ των συλλήψεων ακμαίων αρρένων του *Preys oleae* (πυρηνοτρήτη) σε παγίδες φερομόνης και του εν συνεχεία επιπέδου προσβολής στα άνθη και τους καρπούς της ελιάς σε ελαιώνες της Βοιωτίας. Επίσης, ο Αλεξανδράκης (1987) τεκμηρίωσε τη σχέση μεταξύ αφ’ ενός μεν των συλλήψεων ακμαίων αρρένων των κοκκοειδών *Aonidiella aurantii* και *Planococcus citri* σε παγίδες φερομόνης και αφ’ ετέρου της έντασης (και του χρόνου εμφάνισης) της προσβολής σε εσπεριδοειδώνες του Νομού Χανίων.

Λόγω της ιδιομορφίας του βιολογικού κύκλου και της φαινολογίας των εντόμων, είναι χρήσιμη –αν όχι αναγκαία– η δειγματοληπτική παρακολούθηση της εποχιακής διακύμανσης του πληθυσμού ποιοτικά (διάφορα στάδια) και ποσοτικά. Αυτήν τη μέθοδο ακολούθησε, για διάφορα είδη αρθροπόδων μηδικής στην Κωπαΐδα, λειμώνων στα Ιωάννινα και ακάρεων εδάφους στην Αττική, ομάδα ερευνητών του Γεωπονικού Πανεπιστημίου Αθηνών (Λυκουρέσης κ.ά. 1985, Καπαξίδη κ.ά. 2000, Αντωνάτος κ.ά. 2001, και Ψυχάρης κ.ά. 2001) με στόχο, μεταξύ άλλων, και την αξιολόγηση της ρύπανσης χρησιμοποιώντας τα αρθρόποδα ως βιοδείκτες.

Δεν πρέπει να ξεχνάμε ότι η ζημία μιας καλλιέργειας από κάποιο έντομο συχνά εξαρτάται, εκτός από τον αριθμό των εντόμων, και από άλλους παράγοντες (κλίμα, έδαφος, ηλικία των φυτών κ.τ.λ.), οι οποίοι αλληλεπιδρούν με το έντομο ή το παθογόνο.

Η κινητικότητα των ζώων συχνά εμφανίζει περιοδικότητα εντός του εικοσιτετραώρου. Οι αφίδες π.χ. πετούν κυρίως τις πρωινές ώρες και αργά το απόγευμα. Έτσι έχουν μεγαλύτερες πιθανότητες να συλληφθούν σε παγίδες αναρρόφησης ή κίτρινες κατά τις ώρες αυτές. Παρασιτικά Υμενόπτερα σπάνια πετούν όταν η θερ-

μοκρασία του αέρα είναι χαμηλότερη των 10 °C, ιδίως αν επικρατεί συννεφιά. Η έξοδος των ακμαίων (από τις πούπες) Λεπιδοπτέρων της οικογενείας των Sesiidae (Aegeriidae) λαμβάνει χώρα κάθε ημέρα λίγο μετά την ανατολή του ηλίου και εφόσον η θερμοκρασία του αέρα έχει φθάσει τους 15 °C μέχρι την 10<sup>η</sup> περίπου ώρα (Gorsuch and Karandinos 1974). Τα ακμαία αυτά, μια ώρα περίπου μετά την έξοδό τους, δραστηριοποιούνται για σύζευξη (ελευθερώνοντας φερομόνη τα θήλεα) μέχρι τις πρώτες απογευματινές ώρες και εφόσον η θερμοκρασία του αέρα είναι υψηλότερη των 22 °C (Karandinos 1974). Επομένως, ημέρες που η θερμοκρασία δεν φθάνει τους 22 °C μέχρι αυτές τις ώρες (όπως συμβαίνει συχνά σε μεγάλα γεωγραφικά πλάτη ακόμα και το καλοκαίρι), δεν πρέπει να αναμένουμε συλλήψεις σε παγίδες φερομόνης φύλου.

Ποσοτικές εκτιμήσεις του μεγέθους του πληθυσμού και διαφόρων άλλων παραμέτρων –με δειγματοληψία– είναι αναγκαίες κατ' αρχάς για την *κατανόηση οικολογικών φαινομένων* αλλά και για *πρακτικούς σκοπούς*. Η καταπολέμηση εντόμων που προσβάλλουν τις καλλιέργειες ή τα δασικά δένδρα απαιτεί αποτίμηση της πληθυσμιακής πυκνότητας και της μεταβολής αυτής της πυκνότητας εντός του έτους και μεταξύ ετών. Απαιτεί επίσης την εκτίμηση του ποσοστού παρασιτισμού του εντόμου, τη μέση διάρκεια ζωής του, το αναπαραγωγικό του δυναμικό κ.τ.λ. Οι τιμές αυτές τελικά συσχετίζονται, όπως είδαμε ανωτέρω, με το επίπεδο ζημίας της καλλιέργειας ή του δάσους και με βάση αυτές τις πληροφορίες εκπονείται η *στρατηγική καταπολέμησης*. Τα ανωτέρω ισχύουν και για έντομα που μεταδίδουν ασθένειες στους ανθρώπους και τα ζώα.

Το κουνούπι *Aedes aegypti* είναι ο φορέας μετάδοσης του ιού που προκαλεί τον δάγκειο πυρετό, ο οποίος προσβάλλει κυρίως μικρά παιδιά. Είναι επομένως σοβαρό πρόβλημα δημόσιας υγείας, ιδίως στη Νότια Ασία. Έτσι, από την δεκαετία του '60 ο Διεθνής Οργανισμός Υγείας του ΟΗΕ ίδρυσε στην Bangkok, πρωτεύουσα της Ταϊλάνδης, ένα ερευνητικό Ινστιτούτο του *Aedes* (Begon 1979). Είχε παρατηρηθεί έξαρση της ασθένειας την υγρή περίοδο του έτους με αιχμές κάθε δεύτερο χρόνο. Οι ερευνητές υπέθεσαν ότι αυτοί οι κύκλοι της ασθένειας συσχετίζονταν με τους κύκλους της πληθυσμιακής πυκνότητας των κουνουπιών ή ενδεχομένως με την αποτελεσματικότητά τους ως φορέων της ασθένειας. Το Ινστιτούτο θα ήλεγχε αυτή την υπόθεση με απώτερο στόχο την καταπολέμηση της ασθένειας. Αυτό μπορούσε να επιτευχθεί μόνο εάν η στρατηγική καταπολέμησης εβασίζετο στην κατανόηση της οικολογίας του κουνουπιού. Μεταξύ των εμπλεκόμενων οικολογικών παραμέτρων ήταν η πληθυσμιακή πυκνότητα των κουνουπιών σε διάφορες περιόδους του έτους, η τάση των κουνουπιών να μετακινούνται από τη μία περιοχή στην άλλη και η διάρκεια ζωής των κουνουπιών. Η διάρκεια ζωής μπορεί φυσικά να ποσοτικοποιηθεί μετρώντας τους ρυθμούς επιβίωσης. Αυτές οι παράμετροι έπρεπε να ερευνηθούν υπό φυσικές συνθήκες, αφού έπρεπε να ελεγχθεί η υπόθεση ότι η δυναμική των πληθυσμών

του κουνουπιού αποτελούσε το υπόβαθρο των διακυμάνσεων στην εμφάνιση της ασθένειας. Το κύριο πρόβλημα λοιπόν των ερευνητών ήταν να εκτιμήσουν το μέγεθος και τις άλλες παραμέτρους του πληθυσμού, τα άτομα του οποίου συνεχώς μετακινούνται σε μια μεγάλη έκταση, ενώ συγχρόνως η μια γενιά διαδέχεται την άλλη.

Στη χώρα μας η καταπολέμηση του Δάκου της ελιάς με ψεκασμούς γίνεται –ή πρέπει να γίνεται– όταν εκτιμάται δειγματοληπτικά ότι η πληθυσμιακή πυκνότητα είναι σε τέτοιο ύψος που αναμένεται να προκαλέσει μη ανεκτή απώλεια της παραγωγής. Μετά την ταυτοποίηση και την εργαστηριακή σύνθεσή της, η φερομόνη φύλου του Δάκου (Haniotakis et al. 1977, Mazomenos et al. 1981) χρησιμοποιείται σε παγίδες για την εκτίμηση της πληθυσμιακής πυκνότητας, για καταπολέμηση (μέθοδος μαζικών συλλήψεων των αρρένων) καθώς και για τη μελέτη της βασικής βιολογίας σύζευξης και αναπαραγωγής (Haniotakis et al. 1983, Haniotakis and Pittara 1994). Έτσι διαπιστώθηκε, για παράδειγμα, ότι τα άρρενα ανταποκρίνονται στο ερέθισμα της φερομόνης τις τελευταίες 4 ώρες της φωτόφασης της φωτοπερίόδου 14: 10 (Φ:Σ) και επίσης ότι η ανταπόκριση αυτή μεγιστοποιείται την 6<sup>η</sup> ημέρα της ζωής των ακμαίων. Μελετήθηκε επίσης πειραματικά αλλά και με θεωρητικά υποδείγματα η αποτελεσματικότητα συνδυασμού φερομόνης φύλου και ελκυστικού τροφής (Haniotakis and Vassiliou-Waite 1987, Barclay and Haniotakis 1991). Ο συνδυασμός αυτός χρησιμοποιείται πλέον στην πράξη από τους Έλληνες βιοκαλλιεργητές της ελιάς για την προστασία της παραγωγής τους.

Σοβαρές απώλειες της παραγωγής προκαλούνται επίσης από μύκητες, βακτήρια και ιούς που προσβάλλουν τα φυτά. Ο Διεθνής Οργανισμός Τροφίμων και Γεωργίας σε ετήσιες εκδόσεις του καταγράφει τις απώλειες σε ποσότητες γεωργικών προϊόντων (και τη χρηματική τους αξία) που οφείλονται σε ασθένειες και άλλους εχθρούς των φυτών στις αναπτυσσόμενες και στις αναπτυγμένες χώρες. Σύνοψη τέτοιων στοιχείων παραθέτει ο Τζάμος (2004) στο πρόσφατο βιβλίο του.

Η εκτίμηση της πληθυσμιακής πυκνότητας και των άλλων πληθυσμιακών παραμέτρων δεν είναι αναγκαία μόνο για πληθυσμούς επιβλαβών ειδών αλλά και για πληθυσμούς ειδών που, για κάποιους λόγους, επιθυμούμε να προστατεύσουμε ή να διατηρήσουμε ή γενικά να διαχειριστούμε. Είδη της άγριας πανίδας ή χλωρίδας, είδη κυνηγίου, είδη ψαριών ή άλλων θαλασσιών ειδών ανήκουν σ' αυτές τις κατηγορίες. Ειδικά σήμερα που σε πολλές περιοχές του κόσμου αρκετά είδη βρίσκονται σε κίνδυνο εξάλειψης, οι ποσοτικές οικολογικές μελέτες είναι αναγκαίες για την ανάπτυξη *στρατηγικής διάσωσής τους*. Θα μνημονεύσω τη χαρακτηριστική περίπτωση της θαλάσσιας χελώνας, *Caretta caretta*, την πυκνότητα των «φωλιών» της οποίας –μεταξύ άλλων παραμέτρων– έχει μελετήσει διεξοδικά ο Δημήτρης Μαργαριτούλης και οι συνεργάτες του στη Ζάκυνθο επί σειρά ετών.

Έτσι, σε μια πρόσφατη ενδιαφέρουσα δημοσίευση (Margaritoulis 2005), στην οποία συνοψίζονται τα δεδομένα καταγραφής 19 ετών (1984-2002) των φωλιών της θαλάσσιας χελώνας στη Ζάκυνθο, δίδεται η πυκνότητα –εκφρασμένη ως αριθμός φωλιών ανά τρέχον χιλιόμετρο αμμώδους ακτής– για 6 τοποθεσίες του κόλπου του Λαγανά. Η μέση ετήσια πυκνότητα κυμάνθηκε από 53,7 φωλιές/km στον Ανατολικό Λαγανά έως 1062,8 φωλιές/km στην τοποθεσία Σεκάνια. Η χωρική αυτή διαφοροποίηση της πυκνότητας αποδίδεται κυρίως στη διαφοροποίηση της ανθρώπινης ενόχλησης στις διάφορες περιοχές. Η μέση πυκνότητα φωλιών στις 6 περιοχές κυμάνθηκε από 155,8 φωλιές/km το 1985 έως 366,9 φωλιές/km το 1995. Ο συγγραφέας σωστά παρατηρεί ότι, λόγω της μεγάλης διακύμανσης του αριθμού των φωλιών από έτους εις έτος, δεν είναι εύκολο να αποφανθεί κανείς για τη στασιμότητα ή όχι του πληθυσμού στον χρόνο.

Ο ερευνητής οικολογικών φαινομένων ποσοτικής φύσεως έχει σήμερα στη διάθεσή του πλήθος σύγχρονων τεχνικών και μεθόδων δειγματοληψίας που δεν υπήρχαν παλαιότερα. Για παράδειγμα, τεχνικές υπέρυθρης φωτογράφισης από αεροπλάνα ή δορυφόρους χρησιμοποιούνται σήμερα για την τεκμηρίωση αποφύλλωσης δασικών δένδρων ή την ποσοτική καταγραφή υπέργειων φωλιών (mounds) μυρμηγκιών της φωτιάς (fire ants) (Vogt 2004). Στο άλλο άκρο, τεχνικές με ραδιενεργά ισότοπα ( $^{134}\text{Cs}$ ) χρησιμοποιούνται για τη μελέτη της μετακίνησης θρεπτικών στοιχείων στο έδαφος και στα διάφορα μέρη του φυτού (Arapis 2005).

Επίσης, ο ερευνητής έχει στη διάθεσή του τις μαθηματικές και στατιστικές μεθόδους οι οποίες, σε συνδυασμό με τα προγράμματα ηλεκτρονικών υπολογιστών που έχουν αναπτυχθεί, του επιτρέπουν να επεξεργασθεί πληθώρα ποσοτικών δεδομένων και να αποκαλύψει την πληροφορία που περιέχουν και που τον ενδιαφέρει. Δεν πρέπει να ξεχνούμε ότι το πληθυσμιακό μέγεθος και άλλες παραμέτρους τις μελετούμε με οδηγό τα συγκεκριμένα οικολογικά ερωτήματα που μας ενδιαφέρουν και επίσης ότι το είδος του ερωτήματος καθορίζει και τον *βαθμό αξιοπιστίας* που λογικά θέλουμε να έχει η απάντηση.

Ένα άλλο θέμα λοιπόν που πρέπει να τονίσουμε είναι ότι η εκτίμηση με δειγματοληψία του μεγέθους του πληθυσμού έχει μικρή πρακτική αξία εάν δεν γνωρίζουμε και την αξιοπιστία αυτής της εκτίμησης. Κάθε εκτίμηση του μεγέθους ενός πληθυσμού (ή οποιασδήποτε παραμέτρου) που προκύπτει από δειγματοληψία έχει, ως γνωστόν, κάποιο σφάλμα, με άλλα λόγια έχει ένα βαθμό αξιοπιστίας. Ο βαθμός αυτός γενικά είναι τόσο μεγαλύτερος όσο μεγαλύτερο είναι το *μέγεθος του δείγματος* από το οποίο εκτιμήθηκε η υπό μελέτη παράμετρος.

Στο σημείο αυτό είναι σκόπιμο να διευκρινίσουμε τη διαφορά μεταξύ των όρων *αμερόληπτος* και *αξιόπιστος*. Ο όρος αμερόληπτος (unbiased) χρησιμοποιείται με τη στατιστική έννοια και αναφέρεται σε βασική ιδιότητα του *εκτιμητή*



που χρησιμοποιούμε, ενώ ο όρος αξιοπιστία αναφέρεται στη συγκεκριμένη *εκτίμηση* της παραμέτρου. Για παράδειγμα ο εκτιμητής  $\bar{x}$ , που ορίζεται ως η μέση αριθμητική τιμή μιας παραμέτρου, από τις τιμές ενός τυχαίου δείγματος μεγέθους  $n$  ( $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ), αποδεικνύεται ότι είναι αμερόληπτος εκτιμητής της (άγνωστης) μέσης αριθμητικής τιμής ( $\mu$ ) της παραμέτρου του πληθυσμού ( $E(\bar{x}) = \mu$ ), ανεξάρτητα από το μέγεθος  $n$  του δείγματος. Όμως το μέγεθος του δείγματος επηρεάζει την *αξιοπιστία* της εκτίμησής μας, π.χ. το τυπικό σφάλμα. Είναι θέμα κοινής λογικής ότι η εκτίμηση μιας παραμέτρου που προέκυψε από ένα δείγμα έχει τόσο μεγαλύτερη αξιοπιστία (μικρότερο τυπικό σφάλμα) όσο μεγαλύτερο είναι το μέγεθος του δείγματος. Οι διάφοροι στατιστικοί εκτιμητές που υπάρχουν στη στατιστική βιβλιογραφία χαρακτηρίζονται από *μικρή ή μεγάλη μεροληψία* και από *μικρή ή μεγάλη τυπική απόκλιση* (και επομένως μικρό ή μεγάλο τυπικό σφάλμα). Ο ερευνητής πρέπει να επιλέξει τον «καλύτερο». Συχνά συμβιβάζεται και επιλέγει εκείνον που έχει το μικρότερο «μέσο σφάλμα τετραγώνου» (MSE), που είναι απλά μια συνάρτηση των δύο μεγεθών:  $MSE = (\text{μεροληψία})^2 + \text{διακύμανση}$ . Το θέμα αυτό θα το δούμε λεπτομερώς στο κεφάλαιο 4. Ομαδικός Έλεγχος.

Στην ξένη βιβλιογραφία δειγματοληψίας συχνά χρησιμοποιούνται οι όροι *accuracy* και *precision* (Krebs 1999), που εννοιολογικά έχουν τη σημασία των όρων *αμεροληψία* (unbiased) και *αξιοπιστία* αντίστοιχα και αναφέρονται σε ιδιότητες μετρήσεων που σχετίζονται με την τεχνική της δειγματοληψίας μάλλον παρά με τις στατιστικές ιδιότητες των εκτιμητών των παραμέτρων μιας κατανομής. Εάν π.χ. επανειλημμένες ζυγίσεις ενός αντικειμένου σε έναν ζυγό ακριβείας δώσουν σχεδόν την ίδια τιμή, αλλά η μέση τιμή των ζυγίσεων διαφέρει πολύ από το (γνωστό) πραγματικό βάρος του αντικειμένου (π.χ. λόγω κακής βαθμονόμησης του ζυγού), τότε η ζύγιση έχει μεγάλη αξιοπιστία (precision), αλλά μικρή αμεροληψία (accuracy). Αντίθετα, εάν η μέση τιμή των ζυγίσεων είναι πολύ πλησίον της πραγματικής τιμής του αντικειμένου, αλλά οι επιμέρους ζυγίσεις διαφέρουν αρκετά μεταξύ τους, τότε έχουμε την περίπτωση μεγάλης αμεροληψίας αλλά μικρής αξιοπιστίας των ζυγίσεων.

Στην εκτίμηση των πληθυσμιακών μεγεθών αρθροπόδων εδάφους ή φυλλοστρωμνής, συχνά χρησιμοποιούνται χωνιά Berlesse στα οποία τοποθετείται δείγμα του εδάφους ή της φυλλοστρωμνής. Τα αρθρόποδα του δείγματος απωθούνται από έντονο φως και μετακινούνται στο κάτω μέρος των χωνιών όπου παγιδεύονται σε φιαλίδια με οινόπνευμα. Εάν η ένταση του φωτός είναι μεγάλη και συνεπώς η θερμοκρασία υψηλή, τότε τα μικρότερα αρθρόποδα αφυδατώνονται και πεθαίνουν πριν προφθάσουν να πέσουν στο φιαλίδιο. Έτσι έχουμε μία συστηματική υποεκτίμηση (μεροληπτική) του μεγέθους του πληθυσμού

των μικρών αρθροπόδων, δηλ. η εκτίμηση έχει μικρή αμεροληψία, η οποία δεν βελτιώνεται με πολλές επαναλήψεις.

Αντίθετα, εάν η ένταση του φωτός ή η θερμοκρασία ποικίλλει στον χρόνο ή από δείγμα σε δείγμα, τότε η εκτίμηση του πληθυσμού έχει μικρή αξιοπιστία (μικρή precision), παρόλο που μπορεί να είναι αμερόληπτη. Η αξιοπιστία βελτιώνεται με τις πολλές επαναλήψεις, δηλαδή με το μέγεθος του δείγματος.

Στην εκτίμηση του πληθυσμού μικρών θηλαστικών με τη μέθοδο σύλληψης-σήμανσης-ελευθέρωσης-επανασύλληψης, που θα αναπτύξουμε σε επόμενο κεφάλαιο, είναι δυνατόν τα άτομα που συνελήφθησαν, σημάνθηκαν και απελευθερώθηκαν, να απέκτησαν μια «εμπειρία» που μειώνει την πιθανότητά τους να επανασυλληφθούν («έμαθαν» ν' αποφεύγουν την παγίδευση). Στην περίπτωση αυτή η εκτίμηση του πληθυσμού είναι μεροληπτική (μικρή αμεροληψία) και μάλιστα υπερεκτιμά τον πληθυσμό. Αντίθετα, εάν η τεχνική και η αποτελεσματικότητα της σύλληψης των ζώων διαφέρει από ημέρα σε ημέρα ή εάν το δείγμα είναι μικρό, τότε η αξιοπιστία της εκτίμησης είναι μικρή.

Από τα παραπάνω είναι φανερό ότι ένας από τους βασικούς στόχους της διαδικασίας μιας τυχαίας δειγματοληψίας είναι να επιτύχουμε αξιόπιστη εκτίμηση της υπό μελέτη παραμέτρου. Για να επιτύχουμε τον βαθμό αξιοπιστίας που επιθυμούμε, πρέπει να προσδιορίσουμε τον αριθμό των μετρήσεων –από δυνητικά πολύ μεγάλο και άγνωστο αριθμό– ή, με άλλα λόγια, το μέγεθος του δείγματος που θα εξετάσουμε.

## 2. Το Βέλτιστο Μέγεθος του Δείγματος

Κατ' αρχάς το σύνθημα ερώτημα: ποιο είναι το μέγεθος του δείγματος (αριθμός δειγματοληπτικών μονάδων) που πρέπει να εξετάσουμε, θα πρέπει να διατυπωθεί πιο συγκεκριμένα. Μια τέτοια διατύπωση είναι:

Πόσες δειγματοληπτικές μονάδες πρέπει να εξετάσουμε ούτως ώστε η εκτίμηση της παραμέτρου που ενδιαφέρει να έχει έναν δεδομένο (προκαθορισμένο) βαθμό αξιοπιστίας;

Είναι θέμα κοινής λογικής ότι αν ενδιαφέρομαι να εκτιμήσω το ύψος των δένδρων σε μία συστάδα δάσους και μετρήσω το ύψος ενός μόνο δένδρου, η αξιοπιστία της εκτίμησής μου θα είναι μικρή. Άρα, όσο περισσότερα δένδρα συμπεριλάβω στο δείγμα, τόσο το καλύτερο.

Είναι επίσης θέμα κοινής λογικής ότι, όσο μικρότερες διαφορές ύψους υπάρχουν από δένδρο σε δένδρο της συστάδας, τόσο μικρότερος αριθμός δένδρων επαρκεί για εκτίμηση που θα έχει έναν δεδομένο βαθμό αξιοπιστίας. Στην ακραία περίπτωση π.χ. που τα δένδρα είναι όλα του ίδιου ακριβώς ύψους (όπως οι κο-



λόνες της ΔΕΗ), τότε αρκεί να μετρήσω το ύψος ενός δένδρου και θα έχω την πληροφορία που χρειάζομαι για όλα τα δένδρα. Άρα, όσο η διακύμανση της συγκεκριμένης μεταβλητής στον πληθυσμό είναι μεγαλύτερη, τόσο μεγαλύτερο δείγμα πρέπει να εξετάσω για να έχω την επιθυμητή αξιοπιστία της εκτίμησης.

Ας δούμε τώρα τρεις τρόπους με τους οποίους μπορεί να εκφρασθεί η αξιοπιστία της εκτίμησης και να προσδιοριστεί το αντίστοιχο μέγεθος του δείγματος (Karandinos 1976).

## 2.1. Κριτήριο το Τυπικό Σφάλμα

Έστω η τυχαία μεταβλητή  $x$  (π.χ. ύψος των δένδρων) με μέση τιμή του πληθυσμού  $\mu$  και διακύμανση  $\sigma^2$ . Εάν πάρω ένα δείγμα μεγέθους  $n$  δένδρων και υπολογίσω το μέσο ύψος  $\bar{x}$  των δένδρων του δείγματος ( $\bar{x} = \frac{1}{n} \sum x_i$ ), η διακύμανση της νέας μεταβλητής  $\bar{x}$  είναι κατά τα γνωστά:  $\sigma_{\bar{x}}^2 = \sigma^2/n$ . Το τυπικό σφάλμα είναι  $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ , άρα  $n = \sigma^2/\sigma_{\bar{x}}^2$ .

Είναι ευνόητο ότι όσο μικρότερο είναι το τυπικό σφάλμα  $\sigma_{\bar{x}}$ , τόσο μεγαλύτερη η αξιοπιστία της εκτίμησης  $\bar{x}$ . Το  $\sigma_{\bar{x}}$  είναι λοιπόν ένα μέτρο της αξιοπιστίας. Όμως υπάρχει ένα προφανές πρακτικό πρόβλημα: δεν γνωρίζουμε φυσικά το  $\sigma^2$  του πληθυσμού, και χωρίς αυτή την πληροφορία δεν μπορούμε να υπολογίσουμε το απαιτούμενο μέγεθος του δείγματος στο δειγματοληπτικό μας πρόγραμμα. Το αδιέξοδο αυτό αντιμετωπίζεται χρησιμοποιώντας εκτίμηση  $s_0^2$  από μια προκαταρκτική δειγματοληψία. Αντικαθιστώντας στον ανωτέρω τύπο, έχω την εκτίμηση του απαιτούμενου μεγέθους του δείγματος, δηλαδή:

$$\hat{n} = \frac{s_0^2}{s_{\bar{x}}^2} \quad (1.1)$$

Ο τύπος αυτός μας δίνει το αναγκαίο μέγεθος του δείγματος για να έχει η τελική εκτίμησης μας προκαθορισμένο τυπικό σφάλμα  $s_{\bar{x}}$ .

Το  $s_{\bar{x}}$  και κατ' επέκταση το  $\sigma_{\bar{x}}$  είναι βέβαια ένα μέτρο της αξιοπιστίας της εκτίμησης μας, αλλά όχι πολύ καλό, όπως θα δούμε με ένα απλό παράδειγμα. Μια εκτίμηση πληθυσμιακού μέσου  $\mu = 10$  με  $\sigma_{\bar{x}} = 1,2$  έχει πολύ μικρότερη αξιοπιστία από την εκτίμηση ενός πληθυσμιακού μέσου  $\mu = 100$  με  $\sigma_{\bar{x}} = 1,2$  (με μέγεθος δείγματος  $n$  και στις δύο περιπτώσεις), γιατί στην πρώτη περίπτωση το τυπικό σφάλμα είναι ένα μεγάλο ποσοστό του μέσου, σε σχέση με τη δεύτερη περίπτωση, γεγονός μάλλον ανεπιθύμητο.

## 2.2. Κριτήριο ο Συντελεστής Παραλλακτικότητας

Το ανωτέρω μειονέκτημα του τυπικού σφάλματος (ως μέτρου αξιοπιστίας) αντιμετωπίζεται χρησιμοποιώντας τον Συντελεστή Παραλλακτικότητας. Ο «Συντελεστής Παραλλακτικότητας» του  $\bar{x}$  που χρησιμοποιείται συχνά ως μέτρο αξιοπιστίας είναι:

$$CV(\bar{x}) = \frac{\sigma_{\bar{x}}}{\mu} \quad \text{ή για συντομία} \quad C = \frac{\sigma_{\bar{x}}}{\mu} \quad (1.2)$$

Εκφράζοντας το  $\sigma_{\bar{x}}$  συναρτήσει του  $\sigma$  και  $n$ , η (1.2) γίνεται:

$$C = \frac{\sigma}{\mu\sqrt{n}} \quad (1.3)$$

Λύνοντας ως προς  $n$  έχουμε:

$$n = \frac{\sigma^2}{\mu^2 C^2} \quad (1.4)$$

Χρησιμοποιώντας τις εκτιμήσεις  $\bar{x}_0$  και  $s_0$  των  $\mu$  και  $\sigma$  από προκαταρκτική δειγματοληψία, όπως και στην προηγούμενη περίπτωση, και αντικαθιστώντας στην (1.4) έχω την εκτίμηση  $\hat{n}$  του απαιτούμενου μεγέθους του δείγματος, δηλαδή:

$$\hat{n} = \frac{s_0^2}{\bar{x}_0^2 C^2} \quad (1.5)$$

### Αριθμητικό παράδειγμα 1.1

Έστω ότι από προκαταρκτική δειγματοληψία έχω:  $\bar{x}_0 = 18$  και  $s_0 = 20$ . Επιθυμώ εκτίμηση του μέσου  $\mu$  με συντελεστή παραλλακτικότητας  $C = 0,15$ . Σε έρευνες υπαίθρου συντελεστές παραλλακτικότητας μέχρι και 0,20 είναι γενικά αποδεκτοί. Το απαιτούμενο μέγεθος του δείγματος εκτιμάται:

$$\hat{n} = \frac{400}{(324)(0,0225)} = 55$$

## 2.3. Κριτήριο το Διάστημα Εμπιστοσύνης

Μια τρίτη προσέγγιση στο πρόβλημα υπολογισμού του μεγέθους του δείγματος βασίζεται στα γνωστά από τη στατιστική, σύμφωνα με τα οποία αν πάρουμε ένα τυχαίο δείγμα μεγέθους  $n$  από έναν πληθυσμό που έχει μέση τιμή  $\mu$  και διακύμανση  $\sigma^2$ , τότε, σύμφωνα με το Κεντρικό Οριακό Θεώρημα –Central Limit Theorem– (Hogg and Craig 1965) για επαρκώς μεγάλο δείγμα (περίπου 30 ή μεγαλύτερο) ισχύει η σχέση:

$$\Pr(-Z_{a/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < +Z_{a/2}) = 1 - a$$

ή μετά από ανακατάταξη

$$\Pr(\bar{x} - Z_{a/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{a/2} \frac{\sigma}{\sqrt{n}}) = 1 - a \quad (1.6)$$

όπου  $\bar{x}$  είναι η μέση τιμή του δείγματος,  $(1 - a)$  είναι το επίπεδο σημαντικότητας και  $Z_{a/2}$  είναι το άνω  $a/2$  σημείο της τυπικής κανονικής κατανομής. Υπενθυμίζουμε ότι για  $(1 - a) = 0,95$ ,  $Z_{a/2} = 1,96$ .

Η έκφραση (1.6) διατυπώνει απλά την πρόταση ότι το διάστημα εμπιστοσύνης

$$\bar{x} \pm Z_{a/2} \frac{\sigma}{\sqrt{n}} \quad (1.7)$$

θα περιέχει τον μέσο  $\mu$  με πιθανότητα ίση με  $(1 - a)$ .

Έχουμε λοιπόν τρία στοιχεία: το  $n$ , το  $(1 - a)$  και το εύρος του διαστήματος (1.7).

Δεν έχουμε παρά να αποφασίσουμε για την τιμή των δύο και να λύσουμε ως προς το τρίτο στοιχείο, π.χ. το  $n$ , δηλ. το μέγεθος του δείγματος. Η τιμή του  $(1 - a)$  λαμβάνεται συνήθως στα επίπεδα 0,95 ή 0,99. Σε κάθε περίπτωση φυσικά πρέπει να το αποφασίσουμε.

Το εύρος του διαστήματος εμπιστοσύνης (1.7) μπορεί να οριστεί με τους εξής δύο (τουλάχιστον) τρόπους:

### 2.3.1 Εύρος Διαστήματος Εμπιστοσύνης με βάση σταθερό % ( $D$ ) της παραμέτρου $\mu$

Το εύρος του διαστήματος είναι:

$$Z_{a/2} \frac{\sigma}{\sqrt{n}} = D\mu \quad (1.8)$$

$$\text{Άρα: } n = \frac{(Z_{a/2} \sigma)^2}{(D\mu)^2} \quad \text{και στην πράξη: } \hat{n} = \frac{(Z_{a/2} s_0)^2}{(D\bar{x}_0)^2} \quad (1.9)$$

Αντικαθιστώντας όμως τα  $\mu$  και  $\sigma$ , που δεν είναι γνωστά, με τις προκαταρκτικές τους εκτιμήσεις  $\bar{x}_0$  και  $s_0$ , η κατανομή του  $\frac{\bar{x} - \mu}{s_0 / \sqrt{n}}$  δεν είναι πλέον η τυπική κανονική αλλά η  $t$ -κατανομή. Έτσι στη θέση του  $Z_{a/2}$ , πρέπει να χρησιμοποιήσω  $t_{a/2, n-1}$ , δηλ.

$$\hat{n} = \left( \frac{t_{a/2, n-1} s_0}{D\bar{x}_0} \right)^2 \quad (1.10)$$

Εάν πρόκειται για μέγεθος δείγματος σχετικά μεγάλο (π.χ.  $> 30$ ), τότε οι δύο κατανομές σχεδόν συμπίπτουν και δεν υπάρχει πρόβλημα. Για μικρές όμως

τιμές του  $n$ , για τις οποίες η τιμή  $t_{\alpha/2, n-1}$  στην (1.10) εξαρτάται από το  $n$  (βαθμοί ελευθερίας  $n - 1$ ), η (1.10) πρέπει να λυθεί με τη μέθοδο των επανειλημμένων δοκιμών, όπως θα δούμε στη συνέχεια με ένα παράδειγμα.

### 2.3.2 Εύρος Διαστήματος Εμπιστοσύνης με βάση σταθερό θετικό αριθμό $h$

Επομένως 
$$Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = h .$$

$$\text{Άρα: } n = \left( \frac{Z_{\alpha/2} \sigma}{h} \right)^2 \text{ και στην πράξη: } \hat{n} = \left( \frac{Z_{\alpha/2} s_0}{h} \right)^2 \quad (1.11)$$

Για μικρά  $n$  έχω, όπως προηγουμένως:

$$\hat{n} = \left( \frac{t_{\alpha/2, n-1} s_0}{h} \right)^2 \quad (1.12)$$

#### Αριθμητικό παράδειγμα 1.2

Έστω ότι θέλουμε να εκτιμήσουμε τη μέση τιμή  $\mu$  μιας τυχαίας μεταβλητής  $x$  (π.χ. βάρος), των ατόμων ενός πληθυσμού. Επιθυμούμε η αξιοπιστία να είναι τέτοια ώστε η εκτίμηση να περιέχεται με πιθανότητα 0,95 (δηλ.  $\alpha = 0,05$ ) στο διάστημα εμπιστοσύνης που ορίζεται ως  $\pm 0,10\mu$  (δηλ.  $D = 0,10$  της περίπτωσης 2.3.1 ανωτέρω). Από προκαταρκτική δειγματοληψία έχω:  $\bar{x}_0 = 90$  και  $s_0 = 8$ . Χρησιμοποιώντας τις τιμές αυτές και διάφορες διαδοχικές τιμές του  $n$  στον τύπο (1.10), υπολογίζω το δεύτερο μέρος της παράστασης ούτως ώστε η τιμή της να γίνει ίση με την τιμή  $n$ . Από τις πέντε πρώτες στήλες του πίνακα που ακολουθεί βλέπουμε ότι το μέγεθος του δείγματος  $\hat{n}$  που απαιτείται είναι της τάξεως 5 ή 6.

Έστω ότι με δεδομένη την προκαταρκτική τιμή  $s_0 = 8$  επιθυμούμε η εκτίμηση της μέσης τιμής να περιέχεται με πιθανότητα 0,95 στο διάστημα εμπιστοσύνης

$n$	$n - 1$	$t_{0,025, n-1}$	$(t s_0)^2$	$\frac{(t s_0)^2}{(D \bar{x}_0)^2}$	$\frac{(t s_0)^2}{h^2}$
3	2	4,303	1185,01	14,62	32,92
4	3	3,182	648,01	8,00	18,00
<b>5</b>	4	2,776	493,19	<b>6,09</b>	13,70
<b>6</b>	5	2,571	423,04	<b>5,22</b>	11,75
7	6	2,447	383,22	4,73	10,65
8	7	2,365	357,97	–	9,94
<b>9</b>	8	2,306	340,33	–	<b>9,45</b>
10	9	2,262	327,46	–	9,10

που ορίζεται ως  $\pm 6$  (δηλ.  $h = 6$  στον τύπο 1.12 της περίπτωσης 2.3.2 ανωτέρω). Το μέγεθος του δείγματος προσδιορίζεται και πάλι με διαδοχικές δοκιμές. Από τις στήλες 1-4 και 6 του πίνακα βλέπουμε ότι το μέγεθος του δείγματος που απαιτείται είναι  $\hat{n} = 9$ .

## 2.4. Επιλογή Κατάλληλης Έκφρασης της Αξιοπιστίας

Έστω ότι θέλουμε να εκτιμήσουμε τα μέσα ύψη των δένδρων σε τέσσερις συστάδες Α, Β, Γ, και Δ. Από τις προκαταρκτικές δειγματοληψίες έχουμε εκτιμήσεις για τα μέσα ύψη ( ${}_A\bar{x}_0 = 10$ ,  ${}_B\bar{x}_0 = 10$ ,  ${}_Γ\bar{x}_0 = 4$ , και  ${}_Δ\bar{x}_0 = 12$ , μέτρα), και τις τυπικές αποκλίσεις ( ${}_A S_0 = 2,1$ ,  ${}_B S_0 = 4,7$ ,  ${}_Γ S_0 = 1,41$  και  ${}_Δ S_0 = 1,41$ ). Σχηματικά, οι τέσσερις συστάδες θα έχουν την εμφάνιση του Σχήματος 1.1. Το μέγεθος του δείγματος της δειγματοληψίας θα εξαρτηθεί, όπως είπαμε, από τον τρόπο που θα εκφράσουμε την αξιοπιστία και τον βαθμό που θέλουμε να έχει.

### 2.4.1 Τυπικό Σφάλμα ως μέτρο της αξιοπιστίας

Έστω ότι εκφράζουμε την αξιοπιστία με το *Τυπικό Σφάλμα*, μεγέθους  $s_{\bar{x}} = 0,5$ . Το μέγεθος του δείγματος δίδεται από τον τύπο (1.1).

\* Για τη συστάδα Α, με  ${}_A S_0 = 2,1$  και  ${}_A\bar{x}_0 = 10$ , δίδει:

$${}_A \hat{n} = \frac{(2,1)^2}{(0,5)^2} = 17,64 \text{ δένδρα}$$

\* Για τη συστάδα Β, με  ${}_B S_0 = 4,7$  και  ${}_B\bar{x}_0 = 10$ , δίδει

$${}_B \hat{n} = \frac{(4,7)^2}{(0,5)^2} = 88,36 \text{ δένδρα}$$

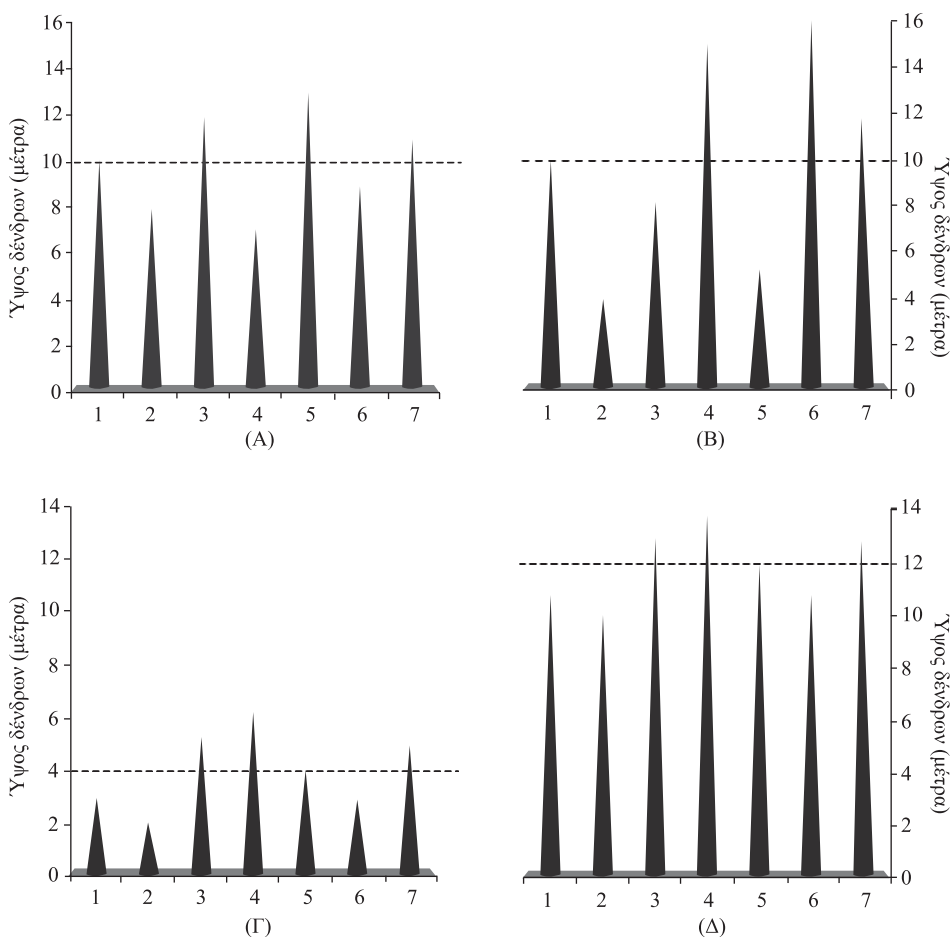
\* Για τη συστάδα Γ, με  ${}_Γ S_0 = 1,41$  και  ${}_Γ\bar{x}_0 = 4$ , δίδει

$${}_Γ \hat{n} = \frac{(1,41)^2}{(0,5)^2} = 7,95 \text{ δένδρα}$$

\* Για τη συστάδα Δ, με  ${}_Δ S_0 = 1,41$  και  ${}_Δ\bar{x}_0 = 12$ , δίδει

$${}_Δ \hat{n} = \frac{(1,41)^2}{(0,5)^2} = 7,95 \text{ δένδρα.}$$

Το μεγαλύτερο δείγμα, που απαιτείται από τη συστάδα Β, ήταν μάλλον αναμενόμενο λόγω μεγάλης διακύμανσης. Όμως, το ίδιο μέγεθος που απαιτείται από τις Γ και Δ δεν ήταν ίσως αναμενόμενο, όπως θα δούμε στη συνέχεια, με άλλη έκφραση της αξιοπιστίας. Οι δεκαδικοί αριθμοί για τα δείγματα έχουν φυσικά μόνο υπολογιστική σκοπιμότητα.



**Σχήμα 1.1.** Σχηματική απεικόνιση τεσσάρων δασοσυστάδων για διερεύνηση του μεγέθους του δείγματος.

#### 2.4.2 Ο Συντελεστής Παραλλακτικότητας ως μέτρο της αξιοπιστίας

Έστω ότι εκφράζουμε την αξιοπιστία με τον *Συντελεστή Παραλλακτικότητας της Εκτίμησης*, μεγέθους  $C = \frac{\sigma_{\bar{x}}}{\mu} = 0,06$  ή 6%. Το μέγεθος του δείγματος δίδεται από τον τύπο (1.5).

\* Για τη συστάδα A, με  ${}_A S_0 = 2,1$  και  ${}_A \bar{x}_0 = 10$ , δίδει:

$${}_A \hat{n} = \frac{(2,1)^2}{(10)^2 (0,06)^2} = 12,25 \text{ δένδρα}$$

\* Για τη συστάδα B, με  ${}_B S_0 = 4,7$  και  ${}_B \bar{x}_0 = 10$ , δίδει:



$${}_B \hat{n} = \frac{(4,7)^2}{(10)^2(0,06)^2} = 61,36 \text{ δένδρα}$$

\* Για τη συστάδα Γ, με  ${}_Γ S_0 = 1,41$  και  ${}_Γ \bar{x}_0 = 4$ , δίδει:

$${}_Γ \hat{n} = \frac{(1,41)^2}{(4)^2(0,06)^2} = 34,51 \text{ δένδρα}$$

\* Για τη συστάδα Δ, με  ${}_Δ S_0 = 1,41$  και  ${}_Δ \bar{x}_0 = 12$ , δίδει:

$${}_Δ \hat{n} = \frac{(1,41)^2}{(12)^2(0,06)^2} = 3,83 \text{ δένδρα.}$$

Βλέπουμε ότι και με την έκφραση αυτή της αξιοπιστίας, η συστάδα Β απαιτεί το μεγαλύτερο δείγμα. Μεταξύ όμως των συστάδων Γ και Δ, η συστάδα Γ απαιτεί πολύ μεγαλύτερο δείγμα, λόγω του μικρού μέσου όρου.

### 2.4.3 Διάστημα Εμπιστοσύνης ως μέτρο της αξιοπιστίας

Έστω ότι εκφράζουμε την αξιοπιστία με το εύρος του Διαστήματος Εμπιστοσύνης της Εκτίμησης και μάλιστα με δύο τρόπους:

(α) Το εύρος του διαστήματος με βάση σταθερό ποσοστό (%)  $D$  της παραμέτρου  $\mu$

Δηλαδή:  $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = D\mu$ . Έστω  $D = 0,05$ . Το μέγεθος του δείγματος δίδει ο τύπος (1.10).

\* Για τη συστάδα Α, με  ${}_Α S_0 = 2,1$  και  ${}_Α \bar{x}_0 = 10$ , δίδει:

$${}_Α \hat{n} = \frac{(1,96)^2(2,1)^2}{(10)^2(0,05)^2} = 67,76 \text{ δένδρα}$$

\* Για τη συστάδα Β, με  ${}_Β S_0 = 4,7$  και  ${}_Β \bar{x}_0 = 10$ , δίδει:

$${}_Β \hat{n} = \frac{(1,96)^2(4,7)^2}{(10)^2(0,05)^2} = 339,44 \text{ δένδρα}$$

\* Για τη συστάδα Γ, με  ${}_Γ S_0 = 1,41$  και  ${}_Γ \bar{x}_0 = 4$ , δίδει:

$${}_Γ \hat{n} = \frac{(1,96)^2(1,41)^2}{(4)^2(0,05)^2} = 190,92 \text{ δένδρα}$$

\* Για τη συστάδα Δ, με  ${}_Δ S_0 = 1,41$  και  ${}_Δ \bar{x}_0 = 12$ , δίδει:

$${}_Δ \hat{n} = \frac{(1,96)^2(1,41)^2}{(12)^2(0,05)^2} = 21,21 \text{ δένδρα.}$$

Και με το κριτήριο αυτό, η συστάδα Β απαιτεί το μεγαλύτερο μέγεθος δείγματος και ακολουθούν κατά σειρά η Γ, η Α και η Δ.

**(β) Το εύρος του διαστήματος με βάση σταθερό θετικό αριθμό  $h$**

Δηλαδή:  $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = h$ . Έστω  $h = 0,8$ . Το μέγεθος του δείγματος δίδει ο τύπος (1.12).

\* Για τη συστάδα Α, με  ${}_A S_0 = 2,1$ , δίδει:

$${}_A \hat{n} = \frac{(1,96)^2 (2,1)^2}{(0,8)^2} = 26,47 \text{ δένδρα}$$

\* Για τη συστάδα Β, με  ${}_B S_0 = 4,7$ , δίδει

$${}_B \hat{n} = \frac{(1,96)^2 (4,7)^2}{(0,8)^2} = 132,59 \text{ δένδρα}$$

\* Για τη συστάδα Γ, με  ${}_Γ S_0 = 1,41$ , δίδει

$${}_Γ \hat{n} = \frac{(1,96)^2 (1,41)^2}{(0,8)^2} = 11,93 \text{ δένδρα}$$

\* Για τη συστάδα Δ, με  ${}_Δ S_0 = 1,41$ , δίδει

$${}_Δ \hat{n} = \frac{(1,96)^2 (1,41)^2}{(0,8)^2} = 11,93 \text{ δένδρα}$$

Παρατηρούμε ότι με το κριτήριο αυτό η κατάταξη των τεσσάρων συστάδων, ως προς το απαιτούμενο μέγεθος του δείγματος, είναι η ίδια με εκείνη του πρώτου κριτηρίου.

**Πίνακας 1.1.** Αναγωγή σε Διάστημα Εμπιστοσύνης της αξιοπιστίας της εκτίμησης με βάση τα υπολογισθέντα μεγέθη του δείγματος.

Παράμετροι Συστάδων	Τυπικό Σφάλμα $S_{\bar{x}} = 0,5$ μέτρα	Συντελ. Παραλ- λακτικότητας $C = 0,06$	Διάστημα Εμπιστοσύνης	
			(α) $D\mu = \pm 0,05\mu$	(β) $h = \pm 0,8$
<b>A</b> ${}_A S_0 = 2,1$ ${}_A \bar{x}_0 = 10$	$\hat{n} = 17,64$ 9,02 – 10,98	$\hat{n} = 12,25$ 8,82 – 11,17	$\hat{n} = 67,76$ 9,5 – 10,5	$\hat{n} = 26,47$ 9,2 – 10,8
<b>B</b> ${}_B S_0 = 4,7$ ${}_B \bar{x}_0 = 10$	$\hat{n} = 88,36$ 9,02 – 10,98	$\hat{n} = 61,36$ 8,82 – 11,17	$\hat{n} = 339,44$ 9,5 – 10,5	$\hat{n} = 132,59$ 9,2 – 10,8
<b>Γ</b> ${}_Γ S_0 = 1,41$ ${}_Γ \bar{x}_0 = 4$	$\hat{n} = 7,95$ 3,02 – 4,98	$\hat{n} = 34,51$ 3,53 – 4,47	$\hat{n} = 190,92$ 3,8 – 4,2	$\hat{n} = 11,93$ 3,2 – 4,8
<b>Δ</b> ${}_Δ S_0 = 1,41$ ${}_Δ \bar{x}_0 = 12$	$\hat{n} = 7,95$ 11,02 – 12,98	$\hat{n} = 3,83$ 10,59 – 13,41	$\hat{n} = 21,21$ 11,4 – 12,6	$\hat{n} = 11,93$ 11,2 – 12,8

Θα είχε ίσως ενδιαφέρον, με τα μεγέθη των δειγμάτων που υπολογίσαμε, να κάνουμε αναγωγή στο ίδιο κριτήριο αξιοπιστίας, π.χ. το διάστημα εμπιστοσύνης της εκτίμησης, και να υπολογίσουμε το εύρος του. Στον Πίνακα 1.1 συνοψίζονται τα αποτελέσματα της διερεύνησης αυτής.

Η επισκόπηση του Πίνακα 1.1 υποδεικνύει ότι από δείγμα μεγέθους 12,2 για τη συστάδα Α –που δίδει Συντελεστή Παραλλακτικότητας 0,06– προκύπτει μάλλον μεγάλο διάστημα (8,82-11,17) εμπιστοσύνης. Επίσης, το ίδιο διάστημα προκύπτει για τη συστάδα Β, αν και το δείγμα είναι αρκετά μεγαλύτερο, προφανώς λόγω της μεγάλης διακύμανσης στον πληθυσμό. Τέλος, για τη συστάδα Γ, το μικρό (7,95) δείγμα το οποίο απαιτείται για τυπικό σφάλμα 0,5 που διερευνήσαμε, σαφώς δίδει μεγάλο σχετικά διάστημα εμπιστοσύνης. Παρόμοιο σχόλιο μπορεί να γίνει και για τον Συντελεστή Παραλλακτικότητας, της συστάδας Δ, για την οποία το πολύ μικρό (3,83) δείγμα δίδει μεγάλο εύρος του διαστήματος εμπιστοσύνης.

### 2.5. Κόστος της Δειγματοληψίας

Μέχρι τώρα απαντήσαμε στο ερώτημα ποιο είναι το ελάχιστο μέγεθος του δείγματος που θα δώσει εκτίμηση της άγνωστης παραμέτρου με τον επιθυμητό βαθμό αξιοπιστίας, την οποία μάλιστα εκφράσαμε με τρεις διαφορετικούς τρόπους.

Υπάρχουν όμως περιπτώσεις που οι διαθέσιμοι πόροι (σε χρήμα, προσωπικό, χρόνο κ.τ.λ.) για τη δειγματοληψία είναι προκαθορισμένοι, δηλ. δεν μπορούν να υπερβούν κάποιο όριο. Τότε γνωρίζοντας το συνολικό διαθέσιμο ποσόν και το κόστος κάθε δειγματοληπτικής μονάδας, υπολογίζουμε τον μέγιστο αριθμό  $n$  δειγματοληπτικών μονάδων που μπορούμε να εξετάσουμε. Με το  $n$  πλέον δεδομένο, λύνουμε τους παραπάνω τύπους (1.1, 1.5, 1.9, 1.11) ως προς τον βαθμό αξιοπιστίας ( $s_{\bar{x}}$ ,  $C$ ,  $D$ ,  $h$ ) που θα έχει η εκτίμησή μας. Εάν η διακύμανση στον πληθυσμό είναι μεγάλη και οι διαθέσιμοι πόροι μικροί, είναι δυνατόν η αξιοπιστία της εκτίμησης να είναι τόσο μικρή που να κρίνουμε ότι δεν έχει νόημα να προχωρήσουμε στη δειγματοληψία. Αντίθετα, αν η αξιοπιστία κριθεί ικανοποιητική, προχωρούμε στη δειγματοληψία γνωρίζοντας όμως ποιος θα είναι περίπου ο βαθμός αξιοπιστίας της εκτίμησης που θα προκύψει. Στις επόμενες σελίδες θα δούμε πώς αντιμετωπίζονται σύνθετα σχήματα δειγματοληψίας, δηλ. σε δύο ή περισσότερα επίπεδα.

### 3. Ειδικές Περιπτώσεις Ασυνεχών Κατανομών

Οι γενικοί τύποι που δώσαμε πιο πάνω για τον υπολογισμό του βέλτιστου μεγέθους του δείγματος και συγκεκριμένα οι τύποι (1.5), (1.9) ή (1.10) και (1.11) ή (1.12), οι

οποίοι αντιστοιχούν στους διαφορετικούς τρόπους ορισμού της αξιοπιστίας της εκτίμησης, μπορούν να χρησιμοποιηθούν για *συνεχείς κατανομές* στις οποίες η μέση τιμή ακολουθεί –για δείγμα μεγαλύτερο του 30– την κανονική κατανομή (κεντρικό οριακό θεώρημα). Εάν η τυχαία μεταβλητή μας είναι *ασυνεχής*, τότε ανάλογα με τη συγκεκριμένη ασυνεχή κατανομή που ακολουθεί, οι αντίστοιχοι γενικοί τύποι που δόθηκαν ήδη μπορούν να πάρουν προσφορότερη μορφή. Θα δούμε αυτές τις μορφές για τις κατανομές: *Διωνυμική*, *Poisson* και *Αρνητική Διωνυμική Κατανομή*, που είναι οι συνηθέστερες σε οικολογικές μελέτες. Η θεωρία που αναπτύχθηκε ήδη και η λογική του υπολογισμού του βέλτιστου μεγέθους του δείγματος που εκθέσαμε παραμένουν αναλλοίωτες.

### 3.1. Διωνυμική Κατανομή

Η Διωνυμική Κατανομή βρίσκει εφαρμογή στις περιπτώσεις που σε κάθε δειγματοληπτική μονάδα σημειώνουμε την *παρουσία* ή *απουσία* κάποιου χαρακτηριστικού, όπως π.χ. προσβολή φυτών από κάποιο έντομο ή ασθένεια.

Εάν πάρουμε τυχαία και εξετάσουμε  $n$  φυτά από έναν πληθυσμό στον οποίο το ποσοστό προσβολής είναι  $p$ , τότε η πιθανότητα τα  $x$  από τα  $n$  φυτά να είναι προσβεβλημένα, δίδεται από τη συνάρτηση:

$$b(n; x, p) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad (1.13)$$

όπου  $x = 0, 1, 2, \dots, n$ .

Η τυχαία μεταβλητή είναι η  $x$ . Από τη μαθηματική στατιστική είναι γνωστό ότι:

$$E(x) = np \quad \text{και} \quad V(x) = np(1-p)$$

όπου  $E(x)$  συμβολίζει τη μαθηματική ελπίδα της  $x$  δηλ. τη μέση τιμή της  $x$ , και  $V(x)$  τη διακύμανση της  $x$ .

Στην πράξη είναι προφανές ότι σε μια δειγματοληψία όπως η παραπάνω, ενδιαφερόμαστε να εκτιμήσουμε την άγνωστη παράμετρο  $p$ , δηλ. το ποσοστό των προσβεβλημένων φυτών. Η εκτίμηση αυτή είναι φυσικά:  $\hat{p} = x/n$ . Εστιάζουμε επομένως την προσοχή μας και αποφασίζουμε τον τρόπο που θα εκφράσουμε την αξιοπιστία της εκτίμησης  $\hat{p}$  και τον βαθμό αξιοπιστίας που επιθυμούμε να έχει κατά τα γνωστά.

Προηγουμένως όμως παρατηρώ ότι:

$$E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{E(x)}{n} = \frac{np}{n} = p \quad (1.14)$$

$$V(\hat{p}) = V\left(\frac{x}{n}\right) = \frac{V(x)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \quad (1.15)$$

Εκφράζουμε τώρα την αξιοπιστία με τους τρεις τρόπους που είδαμε ήδη:

### 3.1.1 Μέτρο αξιοπιστίας ο Συντελεστής Παραλλακτικότητας

Ο αρχικός τύπος:

$$CV(\hat{p}) = \frac{\sigma_{\hat{p}}}{p} \quad (1.16)$$

γίνεται λόγω της (1.15):

$$CV(\hat{p}) = \frac{1}{p} \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(1-p)}{pn}}$$

Λύνοντας ως προς  $n$  και αντικαθιστώντας το  $p$  με την προκαταρκτική εκτίμηση  $\hat{p}_0$ , έχω μια εκτίμηση του ζητούμενου  $n$ :

$$\hat{n} = \frac{(1-\hat{p}_0)}{\hat{p}_0 C^2} \quad (1.17)$$

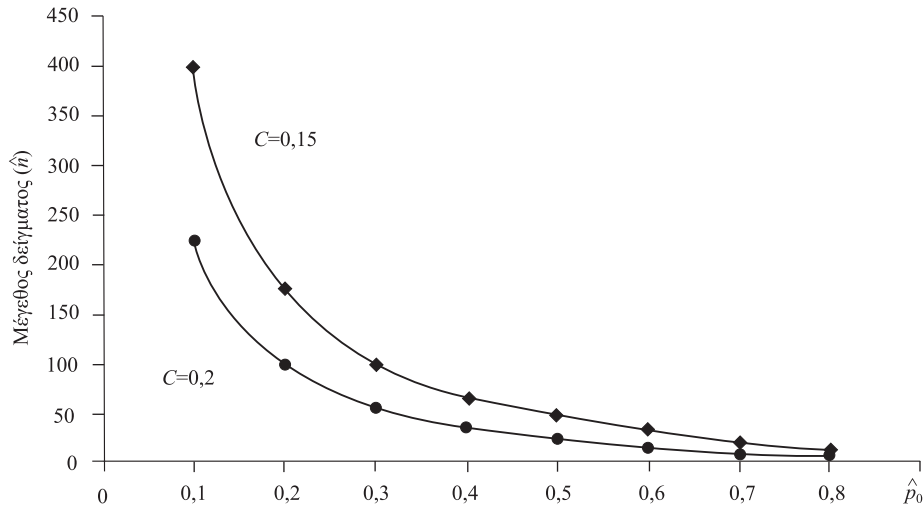
Ο τύπος (1.17) είναι αντίστοιχος του γενικού τύπου (1.5).

#### Αριθμητικό παράδειγμα 1.3

Έστω ότι θέλουμε να εκτιμήσουμε το % των προσβεβλημένων φυτών ενός πληθυσμού και η εκτίμηση να έχει Συντελεστή Παραλλακτικότητας  $C = 0,10$ . Η προκαταρκτική εκτίμηση του ποσοστού είναι  $\hat{p}_0 = 0,30$ . Άρα, εφαρμόζοντας τον τύπο (1.17) έχουμε  $\hat{n} = 233$ . Για διάφορες υποθετικές τιμές της παραμέτρου και διάφορα επιθυμητά  $C$ , τα απαιτούμενα μεγέθη του δείγματος δίδονται στον κατωτέρω πίνακα και στο Σχήμα 1.2.

	$\hat{p}_0$							
	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80
$C = 0,10$	900	400	233	150	100	67	43	25
$C = 0,15$	400	178	104	66	44	30	19	11
$C = 0,20$	225	100	58	37	25	17	11	6
$C = 0,25$	144	64	37	24	16	11	7	4

Τα αποτελέσματα αυτά αποκαλύπτουν ενδιαφέρουσες ιδιότητες: Όπως αναμενόταν, βλέπω ότι όσο μικρότερο  $C$  επιθυμώ τόσο μεγαλύτερο μέγεθος δείγματος πρέπει να εξετάσω. Παρατηρώ επίσης ότι όσο μικρότερο είναι το ποσοστό προσβολής τόσο μεγαλύτερο δείγμα χρειάζομαι για να εκτιμήσω την παράμετρό μου με κριτήριο αξιοπιστίας τον συντελεστή παραλλακτικότητας. Οι σχέσεις αυτές δεν είναι ευθύγραμμες.



Σχήμα 1.2. Μείγεθος δείγματος για διάφορες τιμές της παραμέτρου  $\hat{p}_0$  και επιθυμητές τιμές του Συντελεστή Παραλλακτικότητας  $C$ .

### 3.1.2 Μέτρο αξιοπιστίας το Διάστημα Εμπιστοσύνης

Έστω τώρα ότι επιθυμούμε να ορίσουμε την αξιοπιστία της  $\hat{p}$  με το Διάστημα Εμπιστοσύνης  $Z_{\alpha/2} \sigma_{\hat{p}}$ . Διακρίνουμε δύο υποπεριπτώσεις.

#### (α) Εύρος διαστήματος με βάση σταθερό % ( $D$ ) της παραμέτρου $p$

Έχουμε

$$Z_{\alpha/2} \sigma_{\hat{p}} = Dp \quad (1.18)$$

στην οποία αντικαθιστώντας το  $\sigma_{\hat{p}}$  από την (1.15) έχουμε:

$$Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} = Dp$$

Λύνουμε ως προς  $n$ , αντικαθιστούμε το  $p$  με  $\hat{p}_0$  και έχουμε:

$$\hat{n} = \left( \frac{Z_{\alpha/2}}{D} \right)^2 \cdot \left( \frac{1 - \hat{p}_0}{\hat{p}_0} \right) \quad (1.19)$$

Ο τύπος (1.19) είναι αντίστοιχος του (1.9).

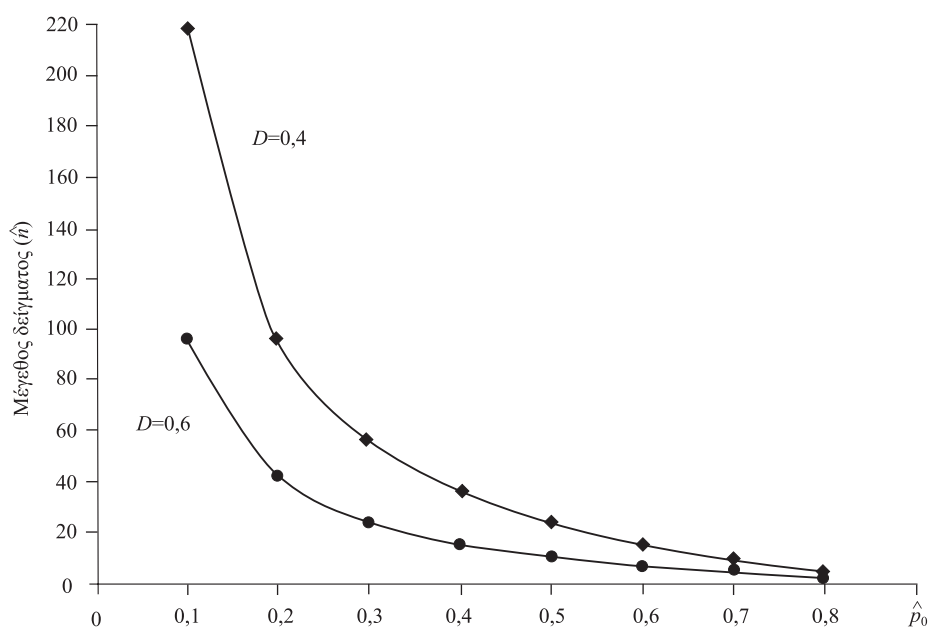
#### Αριθμητικό παράδειγμα 1.4

Εφαρμόζουμε τον τύπο (1.19) με  $Z_{\alpha/2} = 1,96$ ,  $D = 0,4$  και  $\hat{p}_0 = 0,20$  και έχουμε  $\hat{n} = 96$ . Για διάφορες υποτιθέμενες τιμές της παραμέτρου  $\hat{p}_0$  και επιθυμητά  $D$ , τα απαιτούμενα μεγέθη δείγματος δίδονται στον πίνακα που ακολουθεί και στο Σχήμα 1.3.



	$\hat{p}_0$							
	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80
$D = 0,2$	864	384	224	144	96	64	41	24
$D = 0,3$	384	171	99	64	43	28	18	11
$D = 0,4$	216	96	56	36	24	16	10	6
$D = 0,6$	96	43	25	16	11	7	5	3

Και εδώ το μέγεθος του δείγματος συμπεριφέρεται όπως και στην προηγούμενη περίπτωση, αφού η δομή της (1.19) είναι παρόμοια με εκείνη της (1.17).



**Σχήμα 1.3.** Μέγεθος δείγματος για διάφορες τιμές της παραμέτρου  $\hat{p}_0$  της Διωνυμικής Κατανομής και επιθυμητές τιμές του εύρους του Διαστήματος Εμπιστοσύνης ως ποσοστού ( $D$ ) της μέσης τιμής.

**(β) Εύρος Διαστήματος Εμπιστοσύνης με βάση σταθερό αριθμό  $h$**

Έχουμε

$$Z_{\alpha/2} \sigma_{\hat{p}} = h \quad (1.20)$$

από την οποία κατά τον ίδιο τρόπο (τύποι 1.15 και 1.20) βρίσκουμε:

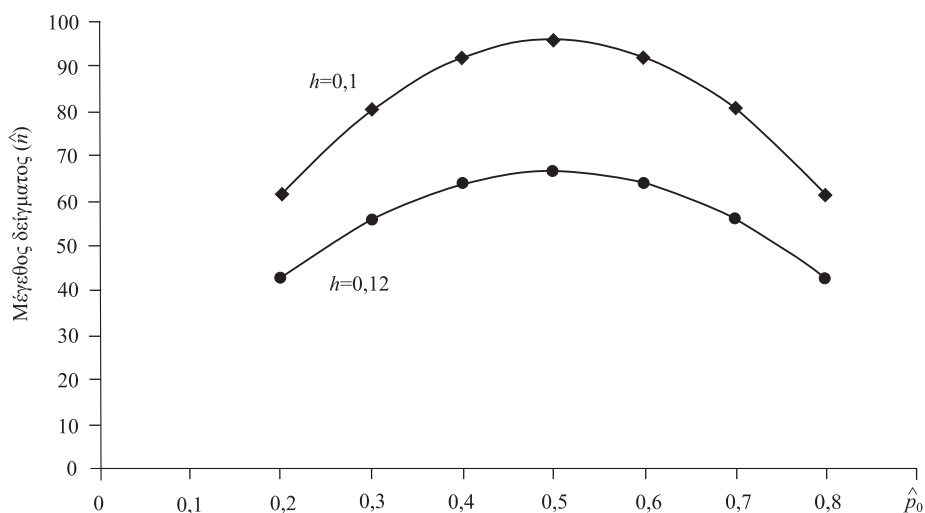
$$\hat{n} = \left( \frac{Z_{\alpha/2}}{h} \right)^2 \hat{p}_0 (1 - \hat{p}_0) \quad (1.21)$$

Ο τύπος (1.21) είναι αντίστοιχος του (1.11) της γενικής περίπτωσης.

*Αριθμητικό παράδειγμα 1.5*

Εφαρμόζουμε τον τύπο (1.21) με  $Z_{\alpha/2} = 1,96$ ,  $h = 0,10$  και  $\hat{p}_0 = 0,20$ . Το  $\hat{n} = 61$ . Για διάφορες υποτιθέμενες τιμές της παραμέτρου  $\hat{p}_0$  και του επιθυμητού  $h$ , έχουμε τα αποτελέσματα στον πίνακα που ακολουθεί και στο Σχήμα 1.4.

	$\hat{p}_0$						
	0,20	0,30	0,40	0,50	0,60	0,70	0,80
$h = 0,08$	96	126	144	150	144	126	96
$h = 0,10$	61	81	92	96	92	81	61
$h = 0,12$	43	56	64	67	64	56	43
$h = 0,14$	31	41	47	49	47	41	31



**Σχήμα 1.4.** Μέγεθος δείγματος για διάφορες τιμές της παραμέτρου  $\hat{p}_0$  της Διωνυμικής Κατανομής και επιθυμητές τιμές  $h$  του εύρους του Διαστήματος Εμπιστοσύνης.

Παρατηρούμε ότι απαιτείται μέγιστος αριθμός δειγμάτων όταν η παράμετρος  $\hat{p}_0$  είναι 0,50. Το αποτέλεσμα αυτό είναι προφανές από τη μορφή της (1.21) η οποία είναι δευτέρου βαθμού συμμετρική συνάρτηση ως προς  $\hat{p}_0$ . Άρα έχει ένα μέγιστο (ή ελάχιστο).

### 3.2. Κατανομή Poisson

Εάν η χωροδιάταξη κάποιων οργανισμών (π.χ. φυτών) ή άλλων γεγονότων στον πληθυσμό των δειγματοληπτικών μονάδων (π.χ. τεμαχίων εδάφους) είναι τυχαία, τότε η τυχαία μεταβλητή  $x$ , ο αριθμός των φυτών στα τεμάχια, έχει την κατανομή του Poisson, δηλ.

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots \quad (1.22)$$

όπου  $\lambda$  η μέση τιμή της μεταβλητής. Αποδεικνύεται ότι:  $E(x) = \lambda$  και  $V(x) = \lambda$ .

Επειδή το  $\lambda$  είναι η μέση τιμή του  $x$  (άρα  $\hat{\lambda} = \frac{1}{n} \sum x_i$ ), η εκτίμηση  $\hat{\lambda}$  έχει διακύμανση:

$$V(\hat{\lambda}) = \frac{V(\sum x_i)}{n^2} = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$$

Ας δούμε τώρα τους τρεις τρόπους ορισμού της αξιοπιστίας.

#### 3.2.1 Μέτρο αξιοπιστίας ο Συντελεστής Παραλλακτικότητας

Έχουμε: 
$$CV(\hat{\lambda}) = \frac{\sigma_{\hat{\lambda}}}{\hat{\lambda}} = \frac{\sqrt{\lambda/n}}{\lambda} = \sqrt{\frac{1}{n\lambda}}$$

Λύνουμε ως προς  $n$  και έχουμε:

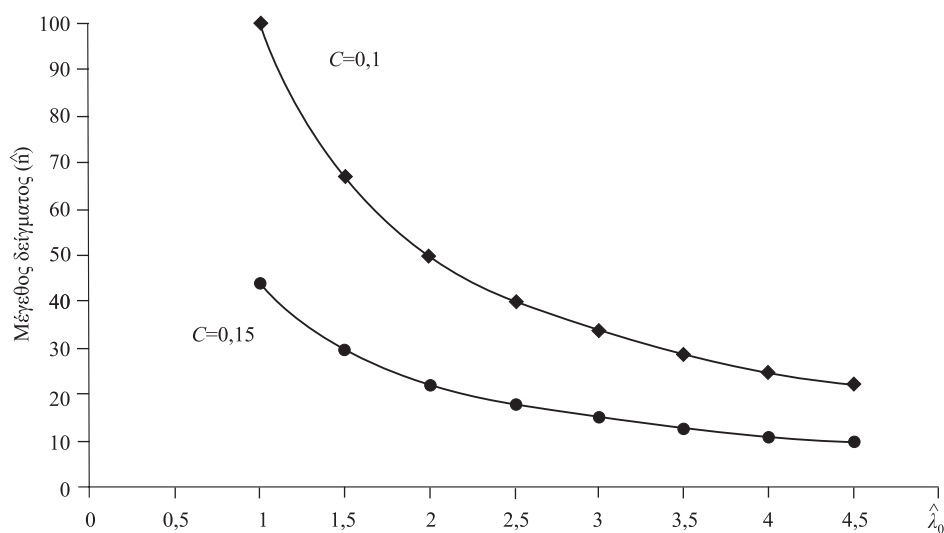
$$\hat{n} = \frac{1}{\hat{\lambda}_0 C^2} \quad (1.23)$$

#### Αριθμητικό παράδειγμα 1.6

Για διάφορες τιμές του  $\hat{\lambda}_0$  και του  $C$  έχω τις κάτωθι εκτιμήσεις του  $n$ :

	$\hat{\lambda}_0$							
	1	1,5	2	2,5	3	3,5	4	4,5
$C = 0,05$	400	267	200	160	133	114	100	89
$C = 0,10$	100	67	50	40	33	29	25	22
$C = 0,15$	44	30	22	18	15	13	11	10
$C = 0,20$	25	17	12	10	8	7	6	5

Παρατηρούμε στον Πίνακα και στο Σχήμα 1.5, ότι, όπως είναι εύλογο, όσο μικρότερο Σ.Π. επιθυμούμε, τόσο μεγαλύτερο δείγμα χρειαζόμαστε. Η σχέση όμως δεν είναι ευθύγραμμη. Παρατηρούμε επίσης ότι όσο μικρότερο είναι το  $\hat{\lambda}_0$ , τόσο μεγαλύτερο δείγμα απαιτείται. Η σχέση και πάλι είναι καμπυλόγραμμη.



**Σχήμα 1.5.** Μέγεθος δείγματος για διάφορες τιμές της παραμέτρου  $\hat{\lambda}_0$  της κατανομής Poisson και επιθυμητές τιμές του Συντελεστή Παραλλακτικότητας  $C$ .

### 3.2.2 Μέτρο αξιοπιστίας το Διάστημα Εμπιστοσύνης

Διακρίνουμε δύο υποπεριπτώσεις:

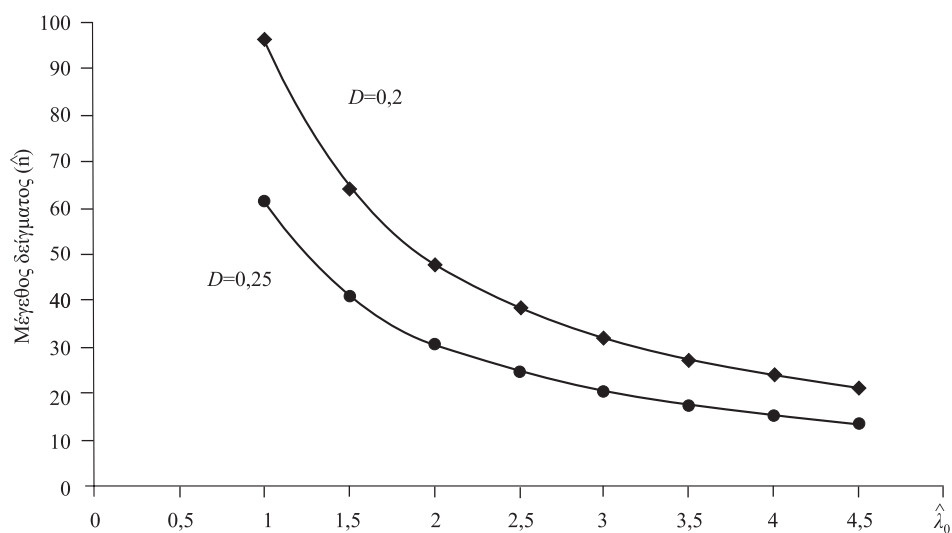
(α) *Εύρος διαστήματος με βάση σταθερό % (D) της παραμέτρου  $\lambda$*

Ο γενικός τύπος (1.9) γίνεται: 
$$\hat{n} = \left( \frac{Z_{\alpha/2}}{D} \right)^2 \frac{1}{\hat{\lambda}_0} \quad (1.24)$$

*Αριθμητικό παράδειγμα 1.7*

Για διάφορες τιμές του  $\hat{\lambda}_0$  και του  $D$  απαιτούνται τα μεγέθη του δείγματος που δίδονται στον κάτωθι Πίνακα και στο Σχήμα 1.6, όταν  $Z_{\alpha/2} = 1,96$ .

	$\hat{\lambda}_0$							
	1	1,5	2	2,5	3	3,5	4	4,5
$D = 0,10$	384	256	192	154	128	110	96	85
$D = 0,15$	171	114	85	68	57	49	43	38
$D = 0,20$	96	64	48	38	32	27	24	21
$D = 0,25$	61	41	31	24	20	17	15	14



Σχήμα 1.6. Μέγεθος δείγματος για διάφορες τιμές της παραμέτρου  $\hat{\lambda}_0$  της κατανομής Poisson και επιθυμητές τιμές του  $D$ .

**(β) Εύρος διαστήματος με βάση σταθερό θετικό αριθμό  $h$**

Ο γενικός τύπος (1.11) γίνεται:

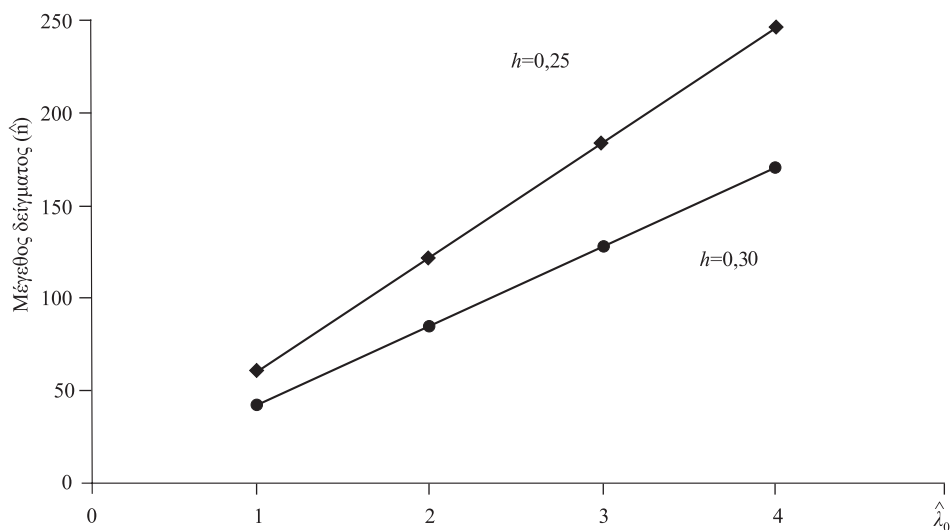
$$\hat{n} = \left( \frac{Z_{\alpha/2}}{h} \right)^2 \hat{\lambda}_0 \quad (1.25)$$

*Αριθμητικό παράδειγμα 1.8*

Για διάφορες τιμές του  $\hat{\lambda}_0$ , του επιθυμητού  $h$  και για  $Z_{\alpha/2} = 1,96$  τα απαιτούμενα μεγέθη του δείγματος δίδονται στον κατωτέρω Πίνακα και στο Σχήμα 1.7:

	$\hat{\lambda}_0$			
	1	2	3	4
$h = 0,15$	171	341	513	683
$h = 0,20$	96	192	288	384
$h = 0,25$	61	123	184	246
$h = 0,30$	43	85	128	171

Είναι προφανές τόσο από τη δομή του τύπου (1.25) όσο και από τα αριθμητικά αποτελέσματα του ανωτέρω πίνακα ότι το μέγεθος του δείγματος είναι ακριβώς ανάλογο του  $\hat{\lambda}_0$  και αντιστρόφως ανάλογο του τετραγώνου του  $h$ .



Σχήμα 1.7. Μέγεθος δείγματος για διάφορες τιμές της παραμέτρου  $\hat{\lambda}_0$  της κατανομής Poisson και επιθυμητές τιμές του  $h$ .

### 3.3. Αρνητική Διωνυμική Κατανομή

Εάν η χωροδιάταξη των οργανισμών στον πληθυσμό των δειγματοληπτικών μονάδων δεν είναι τυχαία αλλά ομαδοποιημένη (δηλ. υπάρχει μεγάλη συγκέντρωση ατόμων σε κάποιες δειγματοληπτικές μονάδες και μικρή σε άλλες), τότε η τυχαία μεταβλητή  $x$ , ο αριθμός των ατόμων στη δειγματοληπτική μονάδα, έχει συχνά την Αρνητική Διωνυμική Κατανομή.

Η κατανομή αυτή έχει δύο παραμέτρους, τη μέση τιμή ( $\mu$ ) της τυχαίας μεταβλητής και τη διακύμανση  $V(x)$ . Συνήθως όμως η διακύμανση της τυχαίας μεταβλητής δίδεται συναρτήσει της παραμέτρου  $\mu$  και μιας άλλης παραμέτρου που επεκράτησε διεθνώς να συμβολίζεται με το γράμμα  $K$ , και η οποία μάλιστα χρησιμοποιείται συχνά στη βιβλιογραφία ως δείκτης του βαθμού ομαδοποίησης της χωροδιάταξης (όσο μεγαλύτερος είναι ο βαθμός ομαδοποίησης, τόσο μικρότερη είναι η τιμή της  $K$ ). Η σχέση που συνδέει τα τρία αυτά μεγέθη είναι:

$$V(x) = \mu + \frac{\mu^2}{K} \quad (1.26)$$

Θα εκφράσουμε λοιπόν το βέλτιστο μέγεθος του δείγματος –με τους τρεις τρόπους που χρησιμοποιήσαμε και στις προηγούμενες κατανομές– συναρτήσει των παραμέτρων  $\mu$  και  $K$ .



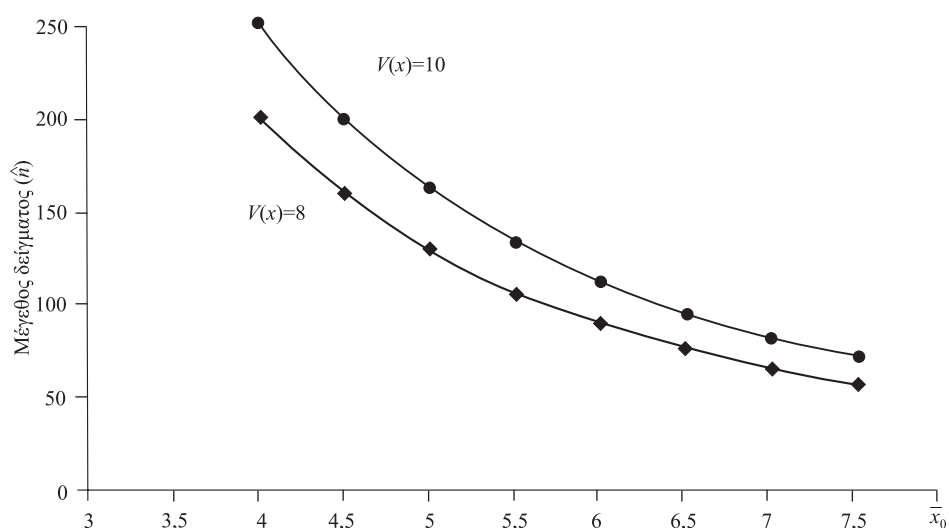
### 3.3.1 Μέτρο αξιοπιστίας ο Συντελεστής Παραλλακτικότητας

Αντικαθιστώντας τη διακύμανση  $\sigma^2$  στην (1.4) από την (1.26) έχουμε:

$$n = \frac{(\mu + K)}{\mu K C^2}$$

και αντικαθιστώντας τη  $\mu$  με την προκαταρκτική εκτίμηση  $\bar{x}_0$  και την  $V(x)$  με την  $\hat{V}_0(x)$  στην (1.26) έχω  $\hat{K}_0$ . Ωστε η εκτίμηση του  $n$  είναι:

$$\hat{n} = \frac{\bar{x}_0 + \hat{K}_0}{\bar{x}_0 \hat{K}_0 C^2} \quad (1.27)$$



**Σχήμα 1.8.** Μέγεθος δείγματος για διάφορες τιμές των παραμέτρων  $\bar{x}_0$  και  $V(x)$ , της Αρνητικής Διωνυμικής Κατανομής (οι οποίες προσδιορίζουν και το  $K$ ) και για  $C = 0,05$ .

Όπως βλέπουμε, το μέγεθος του δείγματος είναι και στην περίπτωση αυτή αντιστρόφως ανάλογο του τετραγώνου του  $C$ . Εξαρτάται επίσης από το  $\bar{x}_0$  και το  $\hat{K}_0$  –με κάποιο σύνθετο τρόπο– όπως βλέπουμε στον τύπο (1.27). Η τιμή του  $\hat{K}_0$  όμως συνδέεται με τη μέση τιμή και τη διακύμανση του  $x$ , όπως φαίνεται στον τύπο (1.26). Έτσι στο αριθμητικό παράδειγμα που ακολουθεί δίδονται οι τιμές του  $\hat{n}$  για διάφορα ζεύγη τιμών του  $\bar{x}_0$  και  $V(x)$ . Εντός παρενθέσεων στο σώμα του πίνακα δίδονται και οι αντίστοιχες τιμές του  $\hat{K}_0$ , ακολουθούμενες από τις τιμές του  $\hat{n}$ .

*Αριθμητικό παράδειγμα 1.9*

Απαιτούμενα μεγέθη δείγματος, από μια Αρνητική Διωνυμική Κατανομή, για διάφορες τιμές του  $\bar{x}_0$  και της διακύμανσης  $V$  και για  $C = 0,05$ . Εντός παρενθέσεων δίδονται οι αντίστοιχες τιμές του  $\hat{K}_0$ .

	$\bar{x}_0$							
	4	5	6	7	8	9	10	11
$V(x) = 8$	(4)	200	(8,33)	128	(18)	89	(49)	65
$V(x) = 10$	(2,66)	250	(5)	160	(9)	111	(16,33)	82
$V(x) = 12$	(2)	300	(3,57)	192	(6)	133	(9,8)	98
$V(x) = 14$	(1,6)	350	(2;78)	224	(4,5)	155	(7)	114

Τα αποτελέσματα του παραδείγματος δίδονται στον Πίνακα και στο Σχήμα 1.8.

Παρατηρούμε στον Πίνακα ότι το μέγεθος του δείγματος αυξάνει γραμμικά με την αύξηση της διακύμανσης. Αντίθετα μειώνεται εκθετικά με την αύξηση της μέσης τιμής της μεταβλητής καθώς και με την αύξηση της τιμής της παραμέτρου  $\hat{K}$ .

**3.3.2 Μέτρο αξιοπιστίας το Διάστημα Εμπιστοσύνης**

Διακρίνουμε πάλι δύο υποπεριπτώσεις:

*(α) Εύρος διαστήματος με βάση σταθερό % (D) της παραμέτρου  $\bar{x}$* 

Αντικαθιστώντας τη διακύμανση στην (1.9) από την (1.26) κατά τα γνωστά έχω:

$$\hat{n} = \frac{Z_{\alpha/2}^2 (\bar{x}_0 + \hat{K}_0)}{\bar{x}_0 \hat{K}_0 D^2} \quad (1.28)$$

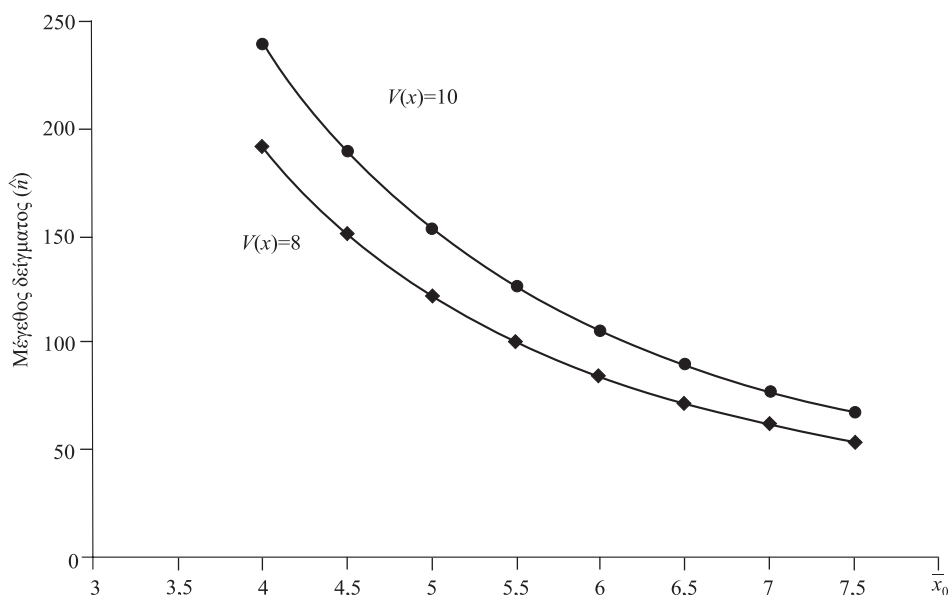
Η δομή του τύπου (1.28) είναι η ίδια με εκείνη του (1.27). Υπολογιστικά διαφέρουν στο ότι στον (1.28) ο αριθμητής πολλαπλασιάζεται με το  $Z_{\alpha/2}^2$ , ο δε παρονομαστής με το  $D^2$  αντί  $C^2$ . Δεν πρέπει βέβαια να μας διαφεύγει η εννοιολογική διαφορά των δύο αυτών συμβόλων.

*Αριθμητικό παράδειγμα 1.10*

Στον πίνακα που ακολουθεί και στο Σχήμα 1.9 δίδεται το μέγεθος του δείγματος για διάφορα ζεύγη τιμών του  $\bar{x}_0$  και της  $V(x)$ . Εντός παρενθέσεων δίδονται οι αντίστοιχες τιμές του  $\hat{K}_0$ . Σε όλες τις περιπτώσεις επιλέξαμε  $D = 0,10$  και  $Z_{\alpha/2} = 1,96$ .

	$\bar{x}_0$							
	4	5	6	7	8	9	10	11
$V(x) = 8$	(4)	192	(8,33)	123	(18)	85	(49)	63
$V(x) = 10$	(2,66)	240	(5)	154	(9)	107	(16,33)	78
$V(x) = 12$	(2)	288	(3,57)	184	(6)	128	(9,8)	94
$V(x) = 14$	(1,6)	336	(2,78)	215	(4,5)	149	(7)	110

Η συμπεριφορά του μεγέθους του δείγματος είναι όπως και στην προηγούμενη περίπτωση.



**Σχήμα 1.9.** Μέγεθος δείγματος για διάφορες τιμές των παραμέτρων  $\bar{x}_0$  και  $V(x)$  της Αρνητικής Διωνυμικής Κατανομής και για  $D = 0,10$ .

**(β) Εύρος διαστήματος με βάση σταθερό θετικό αριθμό  $h$**

Αντικαθιστώντας τη διακύμανση στην (1.11) από την (1.26) έχουμε:

$$\hat{n} = \frac{Z_{\alpha/2}^2 \bar{x}_0 (\hat{K}_0 + \bar{x}_0)}{\hat{K}_0 h^2} \quad (1.29)$$

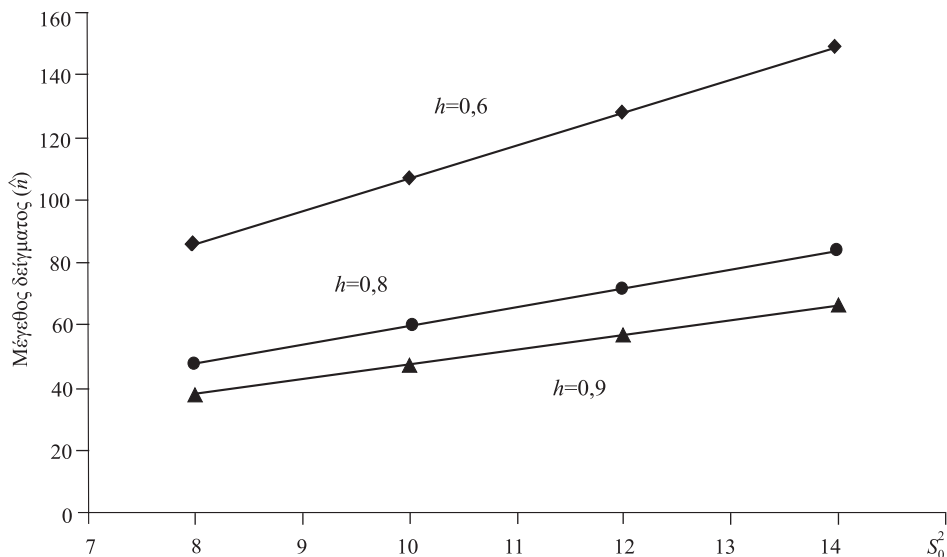
Η δομή του τύπου (1.29) διαφέρει από τη δομή των δύο προηγούμενων όσον αφορά την επίδραση του  $\bar{x}_0$ . Τη μορφή αυτής της επίδρασης θα δούμε σε ένα αριθμητικό παράδειγμα.

### Αριθμητικό παράδειγμα 1.11

Χρησιμοποιώντας τις τιμές  $\bar{x}_0$  και  $V(x)$  των προηγούμενων παραδειγμάτων και τις αντίστοιχες τιμές του  $\hat{K}_0$  για  $Z_{\alpha/2} = 1,96$  και  $h = 0,6$  στον τύπο (1.29) και κάνοντας τους υπολογισμούς παρατηρήσαμε το περίεργο ίσως φαινόμενο ότι το μέγεθος του δείγματος δεν επηρεάζεται από το  $\bar{x}_0$ . Με δεδομένη  $V(x)$  και  $h$  ισχύουν οι ίδιες τιμές του  $\hat{n}$ , για όλα τα  $\bar{x}_0$ . Στον κάτωθι πίνακα και στο Σχήμα 1.10, δίδονται οι τιμές του  $\hat{n}$  για τέσσερις τιμές της  $V(x)$  και τις αντίστοιχες του  $\hat{K}_0$  και για  $\bar{x}_0 = 4$ .

$V(x)$	8	10	12	14
$\hat{K}_0$	4	2,66	2	1,6
$h = 0,4$	192	240	288	336
$h = 0,6$	85	107	128	149
$h = 0,8$	48	60	72	84
$h = 0,9$	38	47	57	66

Ο χαρακτηρισμός «περίεργο» δικαιολογείται γιατί, όπως βλέπουμε, το  $\bar{x}_0$  υπεισέρχεται στον τύπο. Ωριμότερη όμως σκέψη μάς πείθει ότι η συμπεριφορά αυτή είναι αναμενόμενη. Κατ' αρχήν η μαθηματική ερμηνεία προσεγγίζεται αν



**Σχήμα 1.10.** Μέγεθος δείγματος για διάφορα επιθυμητά  $h$ , σταθερή μέση τιμή  $\bar{x}_0 = 4$  και διάφορες τιμές της  $S_0^2$ , (που προσδιορίζουν και την τιμή  $K$ ) της Αρνητικής Διωνυμικής Κατανομής.

λάβουμε υπ' όψη ότι η τιμή του  $\bar{x}_0$  προσδιορίζει –μαζί με την  $V(x)$ – την τιμή του  $\hat{K}_0$ . Υπεισέργεται επομένως και έμμεσα στον τύπο, αλλά προφανώς με τέτοιο τρόπο που εξουδετερώνεται η άμεση παρουσία του. Η δεύτερη πιο πρακτική ερμηνεία προκύπτει αν θυμηθούμε ότι η αξιοπιστία της εκτίμησης, στην περίπτωση που εξετάζουμε, ορίστηκε έτσι ώστε το διάστημα εμπιστοσύνης να έχει κάποιο απόλυτο εύρος ( $h$ ) ανεξάρτητο από τη μέση τιμή της μεταβλητής. Είναι επομένως λογικό το μέγεθος του δείγματος να εξαρτάται από το εύρος αυτό αλλά όχι από τη μέση τιμή της μεταβλητής.

#### 4. Εναλλακτικές Μέθοδοι Εκτίμησης Παραμέτρων

Με τις «εναλλακτικές» αυτές μεθόδους μπορούμε επίσης να εκτιμήσουμε τις τιμές των παραμέτρων της κατανομής μιας τυχαίας μεταβλητής και τα όρια εμπιστοσύνης των εκτιμήσεων αυτών. Οι μέθοδοι αυτές αναπτύχθηκαν από τους στατιστικολόγους πριν από αρκετές δεκαετίες, αλλά μέχρι πρόσφατα δεν είχαν αξιοποιηθεί στην πράξη από οικολόγους και άλλους ερευνητές πεδίου, επειδή η εφαρμογή τους απαιτεί εκτεταμένους υπολογισμούς, που σήμερα πλέον μπορούν να πραγματοποιηθούν με τους Η/Υ και τα διάφορα «πακέτα» που υπάρχουν στην αγορά.

Το μειονέκτημα της υπολογιστικής δυσκολίας των μεθόδων αυτών αντισταθμίζεται από το γεγονός ότι είναι *μη παραμετρικές* μέθοδοι και επομένως δεν είναι απαραίτητο να γνωρίζουμε –ή να θεωρούμε γνωστή– τη στατιστική κατανομή της τυχαίας μεταβλητής και των εκτιμητών των υπό μελέτη παραμέτρων για να υπολογίσουμε τα όρια εμπιστοσύνης των εκτιμήσεών μας. Η έλλειψη αυτής της θεωρητικής γνώσης, αναπληρώνεται «εμπειρικά» με τη διενέργεια μεγάλου αριθμού (εικονικών) δειγματοληψιών από το σώμα των πραγματικών δεδομένων που έχουμε. Η ρηξικέλευθος αυτή προσέγγιση στο πρόβλημα προφανώς υπαγόρευσε και τα ονόματα αυτών των μεθόδων. Η Αγγλική λέξη *Jack-knife* στην καθομιλουμένη σημαίνει *σουγιάς* (με τον οποίο κόβουμε τον γόρδιο δεσμό!). Η λέξη *bootstrap* σημαίνει τις *δερμάτινες θηλιές* που έχουν οι μπότες και διευκολύνουν το φόρεμά τους.

Τις ανωτέρω μεθόδους, με παράθεση συγκεκριμένων αριθμητικών παραδειγμάτων, θα εφαρμόσουμε στην εκτίμηση της *αφθονίας* και της *ποικιλότητας* των ειδών, καθώς επίσης στην εκτίμηση του εύρους και της επικάλυψης των βιοθέσεων στα αντίστοιχα κεφάλαια. Εδώ, είναι ίσως σκόπιμο να παρατηρήσουμε ότι και οι δύο μέθοδοι βασίζονται στην επαναληπτική εκτίμηση των παραμέτρων που μας ενδιαφέρουν χρησιμοποιώντας μέρος των δεδομένων. Η γενική λογική και πορεία εφαρμογής των μεθόδων αυτών έχει ως εξής:

#### 4.1. Η Μέθοδος Jack-knife

Έστω ότι έχουμε  $n$  δείγματα (επαναλήψεις) από έναν πληθυσμό ή βιοκοινότητα. Θα θέλαμε να μάθουμε πόσο καλύτερη θα ήταν η εκτίμησή μας εάν είχαμε ένα ακόμα δείγμα. Αλλά δεν το έχουμε. Έτσι, αντιστρέφουμε την ερώτηση και ζητούμε να μάθουμε πόσο χειρότερη θα ήταν η εκτίμησή μας εάν μας έλλειπε ένα δείγμα. Έχοντας τα  $n$  δείγματα, η *Jack-knife* μέθοδος προχωρεί ως εξής (Krebs 1999):

- Ανασυνδυάζουμε  $n$  φορές τα αρχικά δεδομένα παραλείποντας κάθε φορά μία από τις  $n$  επαναλήψεις.
- Υπολογίζουμε τις λεγόμενες «ψευδοτιμές» της παραμέτρου που μας ενδιαφέρει για κάθε ανασυνδυασμό των δεδομένων:

$$\Phi_i = nS - (n - 1)S_i \quad (1.30)$$

όπου:  $\Phi_i$  = η ψευδοτιμή της *Jack-knife* εκτίμησης  $i$

$n$  = το αρχικό μέγεθος του δείγματος

$S$  = η αρχική στατιστική εκτίμηση της παραμέτρου (με τα  $n$  δείγματα)

$S_i$  = η στατιστική εκτίμηση της παραμέτρου όταν το δείγμα  $i$  είχε παραλειφθεί, ( $i = 1, 2, 3, \dots, n$ ).

Εκτιμούμε τη μέση τιμή και το σταθερό σφάλμα της παραμέτρου που μας ενδιαφέρει από τις  $n$  ψευδοτιμές, δηλαδή:

$$\bar{\Phi} = \frac{1}{n} \sum_{i=1}^n \Phi_i \quad (1.31)$$

$$\text{Σταθερό σφάλμα} = \sqrt{\frac{1}{n(n-1)} \sum_i (\Phi_i - \bar{\Phi})^2} \quad (1.32)$$

Αριθμητικό παράδειγμα εφαρμογής της μεθόδου δίδεται στο κεφάλαιο 5, της Οικολογικής Ποικιλότητας.

#### 4.2. Η Μέθοδος Bootstrap

Η μέθοδος *Bootstrap* απαιτεί μεγαλύτερη υπολογιστική προσπάθεια. Διαφέρει επίσης από τη μέθοδο *Jack-knife* στον τρόπο δειγματοληψίας από το αρχικό σύνολο των δεδομένων. Η μέθοδος *Bootstrap* ερωτά πώς θα ήταν ένα άλλο δείγμα του ίδιου μεγέθους εάν πράγματι είχαμε ένα. Επειδή δεν έχουμε, κάνουμε τυχαία «δειγματοληψία» (με επανατοποθέτηση) από το πραγματικό δείγμα που έχουμε. Επαναλαμβάνουμε αυτήν τη δειγματοληψία πολλές φορές (τουλάχιστο 500) για να εκτιμήσουμε τη μεταβλητότητα της δειγματοληπτικής διαδικασίας.



Από τα δεδομένα αυτών των δειγματοληψιών υπολογίζουμε την παράμετρο που μας ενδιαφέρει και εκτιμούμε τη μέση τιμή και το σταθερό σφάλμα κατά τα γνωστά.

Οι εκτιμήσεις *Bootstrap* είναι μεροληπτικές. Εάν  $\bar{x}_S$  είναι η παρατηρηθείσα μέση τιμή από το αρχικό δείγμα και  $\bar{x}_B$  είναι η *Bootstrap* εκτίμηση της μέσης τιμής, τότε η διορθωμένη –για μεροληψία *Bootstrap*– μέση τιμή θα είναι:  $2\bar{x}_S - \bar{x}_B$ .

## 5. Ιεραρχική Δειγματοληψία με Ίσο Αριθμό Επαναλήψεων

Σε πολλές περιπτώσεις είναι σκόπιμο, εκ των πραγμάτων, να πραγματοποιήσουμε *ιεραρχική δειγματοληψία*, δηλαδή σε υλικό οργανωμένο σε διάφορα (πλέον του ενός) επίπεδα, με ίσο αριθμό επαναλήψεων σε όλα τα επίπεδα. Θα παρουσιάσουμε κατ' αρχάς λεπτομερώς την περίπτωση δύο επιπέδων δειγματοληψίας και στη συνέχεια θα την επεκτείνουμε σε τρία και τέσσερα επίπεδα. Σε κάθε μια περίπτωση διακρίνουμε δύο υποπεριπτώσεις: (α) Μέγιστο ανεκτό κόστος της δειγματοληψίας, (β) Ελάχιστη ανεκτή διακύμανση της εκτίμησης.

### 5.1. Ιεραρχική Δειγματοληψία σε Δύο Επίπεδα

Έστω ότι θέλουμε να εκτιμήσουμε τον αριθμό των προνυμφών κάποιου εντόμου στους καρπούς των δένδρων ενός οπωρώνα. Είναι σκόπιμο να επιλέξουμε τυχαία έναν αριθμό δένδρων και από κάθε δένδρο να επιλέξουμε τυχαία έναν αριθμό καρπών. Στις περιπτώσεις αυτές πρέπει προφανώς να αποφασίσουμε πόσα δένδρα θα συμπεριλάβουμε στο δείγμα μας, και πόσους καρπούς από κάθε δένδρο. Για τις αποφάσεις αυτές, σύμφωνα με όσα έχουμε αναπτύξει στην παρουσίαση της μονοεπίπεδης δειγματοληψίας, σαφώς μας χρειάζεται το κόστος κάθε δειγματοληπτικής μονάδας κάθε επιπέδου και επίσης προκαταρκτική εκτίμηση της διακύμανσης εντός κάθε επιπέδου. Στο ανωτέρω παράδειγμα αντιλαμβανόμαστε εύκολα ότι το κόστος εξέτασης κάθε καρπού, που περιλαμβάνεται στο δείγμα, και της καταμέτρησης –ενδεχομένως στο στερεοσκόπιο– του αριθμού των προνυμφών που «φιλοξενεί», μπορεί να είναι αξιόλογα. Το κόστος όμως που συνδέεται με το κάθε δένδρο που περιλαμβάνεται στο δείγμα, ίσως φανεί αμελητέο. Όμως σε περιπτώσεις που η κάθε δειγματοληπτική μονάδα (δένδρο) πρέπει να καταστραφεί, το κόστος μπορεί να είναι επίσης σημαντικό.

Η πρώτη πληροφορία που χρειαζόμαστε είναι το συνολικό ποσό χρημάτων ( ${}_2C$ ) που έχουμε στη διάθεσή μας για τη δειγματοληψία αυτή. Έστω επίσης ότι το κόστος ανά πρωταρχική δειγματοληπτική μονάδα (δένδρο) είναι  $C_1$  και το κόστος ανά δευτερεύουσα δειγματοληπτική μονάδα (καρπό) είναι  $C_2$ .

Επομένως, εάν το δείγμα μου περιλαμβάνει  $n_1$  δένδρα και  $n_2$  καρπούς ανά δένδρο, τότε συνολικό κόστος  ${}_2C$  της εξέτασης  $n_1$  δένδρων και  $n_1 n_2$  καρπών προφανώς θα είναι:

$${}_2C = C_1 n_1 + C_2 n_1 n_2 \quad (1.33)$$

Η δεύτερη πληροφορία που χρειαζόμαστε είναι οι διακυμάνσεις που αφορούν τις πρωταρχικές και τις δευτερεύουσες δειγματοληπτικές μονάδες, από τις οποίες προσδιορίζεται η διακύμανση της μέσης τιμής προσβολής όλων των καρπών. Οι πληροφορίες αυτές εμπεριέχονται στον πίνακα της γνωστής Ανάλυσης της Παραλλακτικότητας (ANOVA), την οποία πρέπει να κάνουμε, χρησιμοποιώντας τα δεδομένα μιας προκαταρκτικής δειγματοληψίας. Όπως έχουμε ήδη τονίσει, δεν είναι δυνατόν να εκπονήσουμε τη βέλτιστη δειγματοληψία (με όποιο τρόπο και αν την ορίζουμε), εάν δεν έχουμε ήδη κάποιες εκτιμήσεις των διακυμάνσεων στον πληθυσμό που μελετούμε. Τα προκαταρκτικά δεδομένα αναλύουμε όπως δείχνει ο κατωτέρω Πίνακας ANOVA.

Πηγή παραλλακτικότητας	Βαθμοί ελευθερίας β.ε.	Άθροισμα τετραγώνων $SS$	Μέσο τετράγωνο $MS$	Θεωρητική σύσταση του $MS$
Μεταξύ δένδρων	$n_1 - 1$	$SS_\delta$	$MS_\delta = \frac{SS_\delta}{(n_1 - 1)}$	$s^2 = s_2^2 + n_2 s_1^2$
Μεταξύ καρπών (εντός δένδρων)	$n_1(n_2 - 1)$	$SS_\kappa$	$MS_\kappa = \frac{SS_\kappa}{n_1(n_2 - 1)}$	$s_2^2$
Σύνολο	$n_1 n_2 - 1$	$SS_\sigma$		

**Σημείωση:** Στη θεωρητική σύσταση του  $MS$ , οι διακυμάνσεις είναι εκείνες του πληθυσμού που συνήθως συμβολίζονται με το γράμμα  $\sigma^2$ , ενώ οι εκτιμήσεις των συμβολίζονται με  $s^2$ .

Όταν πραγματοποιηθεί η ανάλυση του Πίνακα, τα  $s_2^2$ ,  $s^2$  και  $n_2$  θα είναι γνωστά και έτσι, από τη σχέση  $s^2 = s_2^2 + n_2 s_1^2$ , βρίσκουμε και την τιμή του  $s_1^2$ . Η τιμή  $s^2$  είναι η διακύμανση μεταξύ των τιμών όλων των καρπών. Επειδή το σύνολο των καρπών είναι  $n_1 n_2$ , η διακύμανση της μέσης τιμής  $\bar{x}$  της προσβολής των καρπών είναι, κατά τα γνωστά:

$$s_{\bar{x}}^2 = \frac{s^2}{n_1 n_2} = \frac{s_2^2 + n_2 s_1^2}{n_1 n_2} = \frac{s_2^2}{n_1 n_2} + \frac{s_1^2}{n_1} \quad (1.34)$$

Για την εκπόνηση της βέλτιστης στρατηγικής δειγματοληψίας, μπορούμε να δεχτούμε ως κριτήριο την ελαχιστοποίηση του γινομένου ( ${}_2I$ ) του συνολικού κόστους (1.33) και της διακύμανσης (1.34) της εκτίμησης της μέσης τιμής της

προσβολής. Κάνοντας τις πράξεις, δηλαδή πολλαπλασιάζοντας τις (1.33) και (1.34), βρίσκουμε:

$${}_2\Gamma = {}_2Cs_{\bar{x}}^2 = C_1s_1^2 + C_2s_2^2 + C_2n_2s_1^2 + \frac{C_1s_2^2}{n_2} \quad (1.35)$$

Στη συνάρτηση αυτή παρατηρώ το ενδιαφέρον φαινόμενο ότι το γινόμενο δεν εξαρτάται από το  $n_1$ , τον αριθμό των πρωταρχικών δειγματοληπτικών μονάδων. Με άλλα λόγια, για οποιαδήποτε τιμή του  $n_1$ , υπάρχει μοναδική λύση για το  $n_2$  που ελαχιστοποιεί το ανωτέρω γινόμενο. Για να βρούμε την τιμή αυτή του  $n_2$ , παραγωγίζουμε την  ${}_2\Gamma$ , εξισώνουμε με μηδέν και λύνουμε ως προς  $n_2$  κατά τα γνωστά:

$$\frac{d({}_2\Gamma)}{dn_2} = C_2s_1^2 - \frac{C_1s_2^2}{n_2^2} = 0 \quad (1.36)$$

Άρα ο αριθμός των δευτερευουσών δειγματοληπτικών μονάδων που θα πρέπει να εξετάσουμε στην κανονική (τελική) δειγματοληψία (τον συμβολίζω με  ${}_2N_2$  για να μην γίνεται σύγχυση με το σύμβολο  $n_2$  των μονάδων της προκαταρκτικής δειγματοληψίας), βρίσκεται λύνοντας την (1.36) ως προς  $n_2$ :

$${}_2N_2 = \frac{s_2}{s_1} \sqrt{\frac{C_1}{C_2}} \quad (1.37)$$

Η σχέση αυτή μας αποκαλύπτει ένα άλλο ενδιαφέρον (ίσως και περίεργο) φαινόμενο: ο αριθμός των δευτερευουσών δειγματοληπτικών μονάδων δεν εξαρτάται ούτε από το συνολικό διαθέσιμο χρηματικό ποσό ούτε από τη μέγιστη ανεκτή τυπική απόκλιση της μέσης τιμής. Εξαρτάται όμως και είναι ανάλογος του λόγου των τυπικών αποκλίσεων και μάλιστα, όπως είναι εύλογο, η τυπική απόκλιση στο αντίστοιχο –δεύτερο– επίπεδο, εμφανίζεται στον αριθμητή. Εξαρτάται επίσης και είναι ανάλογος της ρίζας του λόγου των δαπανών και μάλιστα, όπως είναι εύλογο, η δαπάνη ανά δειγματοληπτική μονάδα στο αντίστοιχο επίπεδο εμφανίζεται στον παρανομαστή. Με άλλα λόγια, μεγάλη παραλλακτικότητα και μικρό κόστος, έχουν –όπως είναι εύλογο– ως αποτέλεσμα μεγάλο μέγεθος δείγματος.

Για την ανεύρεση του  ${}_2N_1$  πρέπει να αποφασίσουμε εάν έχουμε περιορισμό του συνολικού κόστους ( ${}_2C$ ) της δειγματοληψίας (Περίπτωση Α) ή εάν θέτουμε τον περιορισμό η διακύμανση της μέσης τιμής ( $s_{\bar{x}}^2$ ) να μην υπερβαίνει μια προαποφασισμένη τιμή (Περίπτωση Β).

### 5.1.1 Μέγιστο ανεκτό κόστος ( ${}_2C$ ) της εκτίμησης. Περίπτωση Α

Έστω ότι το συνολικό ανεκτό κόστος είναι  ${}_2C$ . Επομένως:

$${}_2C = C_1N_1 + C_2N_1N_2 \quad (1.38)$$

Αντικαθιστώ το  $N_2$  στην (1.38) από την (1.37)

$${}_2C = C_1 N_1 + C_2 N_1 \frac{s_2}{s_1} \sqrt{\frac{C_1}{C_2}} = N_1 \left( C_1 + C_2 \frac{s_2}{s_1} \sqrt{\frac{C_1}{C_2}} \right)$$

και λύνω ως προς  $N_1$ :

$${}_2N_1 = \frac{{}_2C}{C_1 + C_2 \frac{s_2}{s_1} \sqrt{\frac{C_1}{C_2}}} = \frac{{}_2C s_1}{\sqrt{C_1} (s_1 \sqrt{C_1} + s_2 \sqrt{C_2})} = \frac{{}_2C s_1}{\sqrt{C_1} \sum_{i=1}^2 s_i \sqrt{C_i}} \quad (1.39)$$

Στη σχέση αυτή παρατηρούμε το εύλογο αποτέλεσμα ότι όσο περισσότερα χρήματα ( ${}_2C$ ) μπορούμε να διαθέσουμε και όσο μεγαλύτερη είναι η τυπική απόκλιση ( $s_1$ ) της τυχαίας μεταβλητής στο εξεταζόμενο (πρώτο) επίπεδο δειγματοληψίας, τόσο μεγαλύτερο δείγμα  ${}_2N_1$  μπορούμε (και πρέπει αντίστοιχα) να εξετάσουμε.

### 5.1.2 Μέγιστη ανεκτή διακύμανση ( $s_x^2$ ) της εκτίμησης. Περίπτωση Β

Έστω ότι δεν θέλουμε η διακύμανση (1.34) της μέσης τιμής να υπερβαίνει μια αποφασισμένη τιμή, που την συμβολίζουμε με  ${}_2V$ . Αντικαθιστώντας στην (1.34) την  $n_2$  από την (1.37) έχουμε:

$${}_2V = \frac{s_2^2}{{}_2N_1 \frac{s_2}{s_1} \sqrt{\frac{C_1}{C_2}}} + \frac{s_1^2}{{}_2N_1} = \frac{1}{{}_2N_1} \left( s_1^2 + s_1 s_2 \sqrt{\frac{C_2}{C_1}} \right)$$

Λύνουμε ως προς  ${}_2N_1$  και έχουμε:

$${}_2N_1 = \frac{s_1^2 + s_1 s_2 \sqrt{\frac{C_2}{C_1}}}{{}_2V} = \frac{s_1 (s_1 \sqrt{C_1} + s_2 \sqrt{C_2})}{{}_2V \sqrt{C_1}} \quad (1.40)$$

Και στη σχέση αυτή παρατηρούμε ότι όσο μικρότερη διακύμανση  ${}_2V$  επιθυμούμε, τόσο μεγαλύτερο δείγμα  ${}_2N_1$  πρέπει να εξετάσουμε.

## 5.2. Ιεραρχική Δειγματοληψία σε Τρία Επίπεδα

Έστω ότι τα επίπεδα είναι: δένδρα, βραχίονες και καρποί. Εξετάζουμε  $n_1$  δένδρα,  $n_2$  βραχίονες ανά δένδρο και  $n_3$  καρπούς ανά βραχίονα.

Έστω επίσης τα αντίστοιχα κόστη:  $C_1, C_2, C_3$  και οι αντίστοιχες διακυμάνσεις:  $s_1^2, s_2^2, s_3^2$ .

Η δομή της ANOVA είναι:

Πηγή παρα- λακτικότητας	Βαθμοί ελευθερίας β.ε.	Άθροισμα τετραγώνων $SS$	Μέσο τετράγωνο $MS$	Θεωρητική σύσταση του $MS$
Μεταξύ δένδρων	$n_1 - 1$	$SS_\delta$	$MS_\delta = \frac{SS_\delta}{n_1 - 1}$	$s_3^2 + n_3 s_2^2 + n_2 n_3 s_1^2$
Μεταξύ βραχιόνων	$n_1(n_2 - 1)$	$SS_\beta$	$MS_\beta = \frac{SS_\beta}{n_1(n_2 - 1)}$	$s_3^2 + n_3 s_2^2$
Μεταξύ καρπών	$n_1 n_2 (n_3 - 1)$	$SS_\kappa$	$MS_\kappa = \frac{SS_\kappa}{n_1 n_2 (n_3 - 1)}$	$s_3^2$
Σύνολο	$n_1 n_2 n_3 - 1$	$SS_\sigma$		

Το συνολικό κόστος, κατά αντιστοιχία της (1.33), θα είναι:

$${}_3C = C_1 n_1 + C_2 n_1 n_2 + C_3 n_1 n_2 n_3 \quad (1.41)$$

Η διακύμανση της μέσης τιμής  $\bar{x}$ , κατά αντιστοιχία της (1.34), θα είναι:

$$s_{\bar{x}}^2 = \frac{s_3^2}{n_1 n_2 n_3} + \frac{s_2^2}{n_1 n_2} + \frac{s_1^2}{n_1} \quad (1.42)$$

Το γινόμενο των (1.41) και (1.42) είναι:

$${}_3\Gamma = {}_3C s_{\bar{x}}^2 = C_1 \left( s_1^2 + \frac{s_2^2}{n_2} + \frac{s_3^2}{n_2 n_3} \right) + C_2 \left( s_2^2 + s_1^2 + \frac{s_3^2}{n_3} \right) + C_3 (s_3^2 + s_2^2 n_3 + s_1^2 n_2 n_3) \quad (1.43)$$

Παρατηρούμε πάλι ότι το γινόμενο δεν εξαρτάται από το  $n_1$ . Εξαρτάται όμως από το  $n_2$  και το  $n_3$ . Επομένως τη συνάρτηση (1.43) θα παραγωγίσουμε δύο φορές, ως προς  $n_2$  και ως προς  $n_3$ , θα τις εξισώσουμε με μηδέν και θα λύσουμε το σύστημα δύο εξισώσεων με δύο αγνώστους.

$$\frac{d({}_3\Gamma)}{dn_2} = -\frac{C_1 s_2^2}{n_2^2} - \frac{C_1 s_3^2}{n_3 n_2^2} + C_2 s_1^2 + C_3 n_3 s_1^2 = 0 \quad (1.44)$$

$$\frac{d({}_3\Gamma)}{dn_3} = -\frac{C_1 s_3^2}{n_2 n_3^2} - \frac{C_2 s_3^2}{n_3^2} + C_3 s_2^2 + C_3 n_2 s_1^2 = 0$$

Η λύση του συστήματος (με λίγη υπομονή) βρίσκεται ότι είναι:

$${}_3N_2 = \frac{s_2}{s_1} \sqrt{\frac{C_1}{C_2}} \quad (1.45)$$

$${}_3N_3 = \frac{s_3}{s_2} \sqrt{\frac{C_2}{C_3}} \quad (1.46)$$

Για να υπολογίσουμε το  ${}_3N_1$  έχουμε πάλι δύο εναλλακτικές περιπτώσεις:

### 5.2.1 Μέγιστο ανεκτό κόστος ( ${}_3C$ ). Περίπτωση Α

Εάν το συνολικό κόστος είναι  ${}_3C$ , τότε στην (1.41) αντικαθιστούμε το  $n_2$  και  $n_3$  με τις αντίστοιχες τιμές από την (1.45) και (1.46) και λύνουμε ως προς  $n_1$ .

$${}_3N_1 = \frac{{}_3C s_1}{\sqrt{C_1} \sum_{i=1}^3 s_i \sqrt{C_i}} \quad (1.47)$$

### 5.2.2 Μέγιστη ανεκτή διακύμανση ( $s_x^2$ ) της εκτίμησης. Περίπτωση Β

Εάν θέλουμε η διακύμανση της μέσης τιμής (1.42) να μην υπερβαίνει μια προ-αποφασισμένη τιμή  ${}_3V$ , τότε αντικαθιστούμε στην (1.42) τις τιμές  $n_2$  και  $n_3$  με τις αντίστοιχες από τις (1.45) και (1.46) και λύνουμε ως προς  $n_1$ .

$${}_3N_1 = \frac{s_1 \sum_{i=1}^3 s_i \sqrt{C_i}}{{}_3V \sqrt{C_1}} \quad (1.48)$$

## 5.3. Ιεραρχική Δειγματοληψία σε Τέσσερα Επίπεδα

Έστω ότι τα επίπεδα είναι: Οπωρώνες, δένδρα, βραχίονες και καρποί. Εξετάζουμε  $n_1$  οπωρώνες,  $n_2$  δένδρα,  $n_3$  βραχίονες και  $n_4$  καρπούς. Οι δειγματοληπτικές μονάδες σε όλα τα επίπεδα πάρθηκαν τυχαία. Έστω επίσης τα αντίστοιχα κόστη:  $C_1, C_2, C_3, C_4$  και οι αντίστοιχες διακυμάνσεις:  $s_1^2, s_2^2, s_3^2, s_4^2$ .

Το συνολικό κόστος φυσικά θα είναι:

$${}_4C = C_1 n_1 + C_2 n_1 n_2 + C_3 n_1 n_2 n_3 + C_4 n_1 n_2 n_3 n_4 \quad (1.49)$$

Η δομή της ANOVA είναι:

Πηγή παρ/τος	Βαθμοί ελευθερίας β.ε.	Άθροισμα τετραγώνων SS	Μέσο τετράγωνο MS	Θεωρητική σύσταση του MS
Μεταξύ οπω/νων	$n_1 - 1$	$SS_o$	$MS_o = \frac{SS_o}{\beta.ε.}$	$s_4^2 + n_4 s_3^2 + n_4 n_3 s_2^2 + n_4 n_3 n_2 s_1^2$
Μεταξύ δένδρων	$n_1(n_2 - 1)$	$SS_\delta$	$MS_\delta = \frac{SS_\delta}{\beta.ε.}$	$s_4^2 + n_4 s_3^2 + n_4 n_3 s_2^2$
Μεταξύ βρα/νων	$n_1 n_2 (n_3 - 1)$	$SS_\beta$	$MS_\beta = \frac{SS_\beta}{\beta.ε.}$	$s_4^2 + n_4 s_3^2$
Μεταξύ καρπών	$n_1 n_2 n_3 (n_4 - 1)$	$SS_\kappa$	$MS_\kappa = \frac{SS_\kappa}{\beta.ε.}$	$s_4^2$
Σύνολο	$n_1 n_2 n_3 n_4 - 1$	$SS_\sigma$		

Η διακύμανση της μέσης τιμής  $\bar{x}$  θα είναι:

$$\begin{aligned}
 s_{\bar{x}}^2 &= \frac{s_4^2 + n_4 s_3^2 + n_4 n_3 s_2^2 + n_4 n_3 n_2 s_1^2}{n_1 n_2 n_3 n_4} \\
 &= \frac{s_4^2}{n_1 n_2 n_3 n_4} + \frac{s_3^2}{n_1 n_2 n_3} + \frac{s_2^2}{n_1 n_2} + \frac{s_1^2}{n_1}
 \end{aligned}
 \tag{1.50}$$

Εργαζόμαστε όπως και προηγουμένως: Πολλαπλασιάζουμε την (1.49) επί την (1.50). Στο γινόμενο το  $n_1$  έχει εξαλειφθεί. Παραγωγίζουμε το γινόμενο αυτό ( $_4\Gamma$ ) ως προς  $n_2$ ,  $n_3$  και  $n_4$ . Θα έχουμε:

$${}_4N_2 = \frac{s_2}{s_1} \sqrt{\frac{C_1}{C_2}} \tag{1.51}$$

$${}_4N_3 = \frac{s_3}{s_2} \sqrt{\frac{C_2}{C_3}} \tag{1.52}$$

$${}_4N_4 = \frac{s_4}{s_3} \sqrt{\frac{C_3}{C_4}} \tag{1.53}$$

Για τον υπολογισμό του  ${}_4N_1$ , έχουμε πάλι δύο εναλλακτικές περιπτώσεις:

### 5.3.1 Μέγιστο ανεκτό κόστος ( ${}_4C$ ). Περίπτωση Α

Εάν το συνολικό κόστος  ${}_4C$  είναι δεδομένο, τότε στην (1.49) αντικαθιστούμε τα  $n_2, n_3$  και  $n_4$  με τις αντίστοιχες τιμές από τις (1.51), (1.52) και (1.53) και λύνουμε ως προς  $n_1$ . Θα βρούμε:

$${}_4N_1 = \frac{{}_4Cs_1}{\sqrt{C_1} \sum_{i=1}^4 s_i \sqrt{C_i}} \quad (1.54)$$

### 5.3.2 Μέγιστη ανεκτή διακύμανση ( $s_{\bar{x}}^2$ ) της εκτίμησης. Περίπτωση Β

Εάν θέλουμε η διακύμανση της μέσης τιμής (1.50) να μην υπερβαίνει μια προ-αποφασισμένη τιμή  ${}_4V$ , τότε αντικαθιστούμε στην (1.50) τις τιμές  $n_2, n_3$  και  $n_4$  με τις αντίστοιχες από τις (1.51), (1.52) και (1.53) και λύνουμε ως προς  $n_1$ :

$${}_4N_1 = \frac{s_1 \sum_{i=1}^4 s_i \sqrt{C_i}}{{}_4V \sqrt{C_1}} \quad (1.55)$$

Προσέξτε τις συμμετρίες που προσδίδουν μια ξεχωριστή ομορφιά στις μαθηματικές αυτές σχέσεις. Πολύ εύκολα μπορείτε εάν θέλετε να βρείτε τις αντίστοιχες λύσεις και στην περίπτωση που έχετε πέντε ή και περισσότερα επίπεδα δειγματοληψίας, αν και σπανίως εμφανίζονται τέτοιες περιπτώσεις στην πράξη.

#### Αριθμητικό παράδειγμα 1.12

Προτιθέμεθα να πραγματοποιήσουμε μια δειγματοληψία δύο επιπέδων (δένδρα και καρπούς) για να εκτιμήσουμε τον αριθμό των κοκκοειδών ανά καρπό σε ένα μεγάλο οπωρώνα. Έστω ότι το κόστος ανά δειγματοληπτική μονάδα του πρώτου επιπέδου (δένδρα) είναι  $C_1 = 20$  χρηματικές μονάδες και του δεύτερου επιπέδου (καρποί) είναι  $C_2 = 2$  μονάδες:

Δένδρα, $1, \dots, n_1$	Καρποί, $1, \dots, n_2$						Σύνολο ανά δένδρο $X_{i\cdot}$	Γενικό Σύνολο $X_{\cdot\cdot}$
	1	2	3	4	5	6		
1	5	6	2	9	2	1	25	
2	2	4	8	12	6	4	36	
3	4	7	10	9	13	11	54	
4	6	16	10	8	10	11	61	176



Η προκαταρκτική δειγματοληψία  $n_1 = 4$  δένδρων και  $n_2 = 6$  καρπών ανά δένδρο (η οποία προφανώς κόστισε  $20 \cdot 4 + 2 \cdot 4 \cdot 6 = 128$  μονάδες) έδωσε τα δεδομένα που παρουσιάζονται στον ανωτέρω Πίνακα. Η μέση τιμή είναι  $176/24 = 7,3333$  κοκκοειδή ανά καρπό.

Όπως γνωρίζουμε από τα στοιχεία στατιστικής ανάλυσης, για να πραγματοποιήσουμε την ανάλυση της παραλλακτικότητας, κάνουμε τις παρακάτω πράξεις:

$$\text{«Διορθωτικός» παράγων: } D = \frac{(X_{..})^2}{n_1 n_2} = \frac{(176)^2}{(4)(6)} = 1290,67$$

Συνολικό Άθροισμα Τετραγώνων:

$$\sum_1^{24} X_{ij}^2 - D = (5^2 + 6^2 + \dots + 10^2 + 11^2) - 1290,67 = 1642 - 1290,67 = 351,33$$

Μεταξύ Δένδρων Άθροισμα Τετραγώνων:

$$\sum_1^4 X_{i.}^2 - D = \frac{(25^2 + 36^2 + 54^2 + 61^2)}{6} - 1290,67 = 1426,33 - 1290,67 = 135,66$$

Μεταξύ Καρπών στο Δένδρο Άθροισμα Τετραγώνων =  $351,33 - 135,66 = 215,66$ .

Με τα στοιχεία αυτά, ο Πίνακας της Ανάλυσης της Παραλλακτικότητας είναι:

Πηγή παραλλακτικότητας	Βαθμοί ελευθερίας	Άθροισμα τετραγώνων $SS$	Μέσο τετράγωνο $MS$	Θεωρητική σύσταση του $MS$
Μεταξύ δένδρων	$n_1 - 1 = 3$	135,67	45,223	$s^2 = s_2^2 + 6s_1^2$
Μεταξύ καρπών	$n_1(n_2 - 1) = 20$	215,66	10,783	$s_2^2$
Σύνολο	$n_1 n_2 - 1 = 23$	351,33		

Επομένως, από τα στοιχεία αυτά έχουμε την προκαταρκτική εκτίμηση  $s_2^2 = 10,783$  και  $s^2 = 45,223$ . Άρα λύνοντας την  $45,223 = 10,783 + 6s_1^2$  ως προς  $s_1^2$  βρίσκουμε  $s_1^2 = 5,74$ . Από την προκαταρκτική αυτή δειγματοληψία, η εκτίμηση της διακύμανσης της μέσης τιμής είναι, σύμφωνα με την (1.34):  $s_{\bar{x}}^2 = 45,223/24 = 1,8843$  και  $s_{\bar{x}} = 1,3727$ . Έχουμε επίσης:  $s_1 = 2,396$ ,  $s_2 = 3,284$  και  $s = 6,725$ .

Με τα στοιχεία αυτά υπολογίζουμε τον αριθμό των δειγματοληπτικών μονάδων δευτέρου επιπέδου που πρέπει να εξετάσουμε για να ελαχιστοποιήσουμε το γινόμενο κόστους και διακύμανσης της εκτίμησης της μέσης τιμής των κοκκοειδών ανά καρπό, χρησιμοποιώντας την παράσταση (1.37):

$${}_2 N_2 = \frac{3,284}{2,396} \sqrt{\frac{20}{2}} = 4,334 \text{ καρπούς ανά δένδρο.}$$

Για να υπολογίσουμε και τον αριθμό των δένδρων, διακρίνουμε τις δύο περιπτώσεις που αναπτύξαμε.

**Περίπτωση Α.** Έστω ότι το συνολικό κόστος είναι  ${}_2C = 600$  μονάδες. Χρησιμοποιώντας την (1.39) έχουμε:

$${}_2N_1 = \frac{(600)(2,396)}{\sqrt{20} [2,396\sqrt{20} + 3,284\sqrt{2}]} = 20,9 \text{ δένδρα.}$$

Σαφώς, η κατανομή των χρημάτων στην προκαταρκτική δειγματοληψία στα δύο επίπεδα δεν ήταν η βέλτιστη, αφού βλέπουμε ότι χρειάζονται περισσότερα δένδρα (21 αντί 4) και λιγότεροι καρποί (4 αντί 6). Με την ανωτέρω κατανομή και τις 600 χρηματικές μονάδες, η διακύμανση της μέσης τιμής –τύπος (1.34)– θα είναι:

$$s_{\bar{x}}^2 = \frac{10,783}{(20,9)(4,334)} + \frac{5,74}{20,9} = 0,3936$$

και το τυπικό σφάλμα  $s_{\bar{x}} = \sqrt{0,3936} = 0,6274$ , δηλαδή πολύ μικρότερο του 1,3727, που ίσχυε στην προκαταρκτική δειγματοληψία.

**Περίπτωση Β.** Έστω ότι η διακύμανση  $s_{\bar{x}}^2$  της μέσης τιμής δεν θέλουμε να υπερβαίνει την τιμή 0,81 και άρα το τυπικό σφάλμα την τιμή  $s_{\bar{x}} = 0,9$ . Σύμφωνα με την (1.40), έχουμε:

$${}_2N_1 = \frac{(2,396)(2,396\sqrt{20} + 3,284\sqrt{2})}{0,81\sqrt{20}} = 10,15 \text{ δένδρα}$$

Η δειγματοληψία αυτή θα κοστίσει  $(20)(10,15) + (2)(10,15)(4,334) = 290,98$  μονάδες. Στην πράξη θα εξετάσουμε φυσικά ακέραιες δειγματοληπτικές μονάδες, π.χ. 5 καρπούς ανά δένδρο και 11 δένδρα. Η δειγματοληψία αυτή θα κοστίσει 330 χρηματικές μονάδες και θα δώσει εκτίμηση της τυπικής απόκλισης της μέσης τιμής ακόμα μικρότερη της τιμής 0,9, η οποία είναι ήδη αρκετά μικρότερη εκείνης (1,3727) από την προκαταρκτική δειγματοληψία.

### Αριθμητικό παράδειγμα 1.13

Ας δούμε τώρα την εφαρμογή της μεθόδου σε δειγματοληψία *τριών επιπέδων*. Θα χρησιμοποιήσουμε πραγματικά δεδομένα (Snedecor and Cochran 1967) χημικών αναλύσεων για τον προσδιορισμό της συγκέντρωσης ασβεστίου (% ξηράς ουσίας) σε φύλλα του φυτού *Brassica rapa*.

Ελήφθησαν τυχαία  $n_1 = 4$  φυτά, από κάθε φυτό ελήφθησαν  $n_2 = 3$  φύλλα και σε κάθε φύλλο έγιναν  $n_3 = 2$  χημικοί προσδιορισμοί της συγκέντρωσης του ασβεστίου. Τα δεδομένα των 24 προσδιορισμών δίδονται στον Πίνακα που ακολουθεί.

Η μέση τιμή του % περιεκτικότητας σε ασβέστιο είναι:  $\bar{X} = 72,29/24 = 3,0121$ .

Από τα δεδομένα του Πίνακα υπολογίζουμε επίσης τα στοιχεία που χρειαζόμαστε για την Ανάλυση της Παραλλακτικότητας:

Φυτό, $i$ $i = 1 \dots n_1$	Φύλλα, $j$ $j = 1 \dots n_2$	Προσδιορισμοί, $X_{ijk}$		$X_{ij\bullet}$	$X_{i\bullet\bullet}$	$X_{\bullet\bullet\bullet}$
1	1	3,28	3,09	6,37	19,05	72,29
	2	3,52	3,48	7,00		
	3	2,88	2,80	5,68		
2	1	2,46	2,44	4,90	13,07	
	2	1,87	1,92	3,79		
	3	2,19	2,19	4,38		
3	1	2,77	2,66	5,43	17,71	
	2	3,74	3,44	7,18		
	3	2,55	2,55	5,10		
4	1	3,78	3,87	7,65	22,46	
	2	4,07	4,12	8,19		
	3	3,31	3,31	6,62		

«Διορθωτικός» παράγων:  $C = \frac{(X_{\bullet\bullet\bullet})^2}{n_1 n_2 n_3} = \frac{(72,29)^2}{24} = 217,7435$ .

Προσδιορισμοί:  $\sum_1^{24} X_{ijk}^2 - C = 3,28^2 + \dots + 3,31^2 - 217,7435 = 10,2704$ .

Φύλλα:  $\frac{1}{n_3} \sum_1^{12} X_{ij\bullet}^2 - C = \frac{(6,37^2 + \dots + 6,62^2)}{2} - 217,7435 = 10,1905$ .

Φυτά:  $\frac{1}{n_2 n_3} \sum_1^4 X_{i\bullet\bullet}^2 - C = \frac{(19,05^2 + \dots + 22,46^2)}{6} - 217,7435 = 7,5603$ .

Φύλλα στο φυτό = Φύλλα - Φυτά =  $10,1905 - 7,5603 = 2,6302$ .

Προσδιορισμοί στο φύλλο = Προσδιορισμοί - Φύλλα =  $10,2704 - 10,1905 = 0,0799$ .

Με τα στοιχεία αυτά συμπληρώνουμε στον Πίνακα Ανάλυσης της Παραλλακτικότητας:

Πηγή παραλλακτικότητας	Βαθμοί ελευθερίας	Άθροισμα τετραγώνων $SS$	Μέσο τετράγωνο $MS$	Θεωρητική σύνθεση του $MS$
Μεταξύ φυτών	$n_1 - 1 = 3$	7,5603	2,5201	$s_3^2 + n_3 s_2^2 + n_2 n_3 s_1^2$
Φύλλα στα φυτά	$n_1(n_2 - 1) = 8$	2,6302	0,3288	$s_3^2 + n_3 s_2^2$
Προσδιορισμοί στα φύλλα	$n_1 n_2(n_3 - 1) = 12$	0,0799	0,0067	$s_3^2$
Σύνολο	$n_1 n_2 n_3 - 1 = 23$	10,2704		

Επομένως από τα στοιχεία αυτά έχουμε:

$$s_3^2 = 0,0067, \quad s_2^2 = \frac{0,3288 - 0,0067}{2} = 0,1610$$

και 
$$s_1^2 = \frac{2,5201 - 0,3288}{6} = 0,3652$$

Επίσης  $s^2 = 2,5201$  άρα  $s_{\bar{x}}^2 = \frac{2,5201}{24} = 0,1050$ ,  $s_{\bar{x}} = 0,3240$

Έστω ότι τα κόστη είναι:  $C_1 = 10$ ,  $C_2 = 2$  και  $C_3 = 40$  χρηματικές μονάδες.

Μπορούμε τώρα να υπολογίσουμε τον αριθμό των δειγματοληπτικών μονάδων σε κάθε επίπεδο. Χρησιμοποιώντας τις (1.45) και (1.46) έχουμε:

$${}_3N_2 = \frac{0,4012}{0,6043} \sqrt{\frac{10}{2}} = 1,48 \quad \text{φύλλα, δηλαδή στην πράξη 2 φύλλα.}$$

$${}_3N_3 = \frac{0,0818}{0,4012} \sqrt{\frac{2}{40}} = 0,0456 \quad \text{προσδιορισμούς ανά φύλλο. Προφανώς στην πρά-$$

ξη θα κάνουμε έναν προσδιορισμό ανά φύλλο.

**Περίπτωση Α.** Έστω ότι τα διαθέσιμα χρήματα είναι 3000 μονάδες. Η (1.47) γίνεται:

$$\begin{aligned} {}_3N_1 &= \frac{(3000)(0,6043)}{\sqrt{10}(0,6043\sqrt{10} + 0,4012\sqrt{2} + 0,0818\sqrt{40})} \\ &= \frac{1812,9}{(3,1623)(2,9956)} = 191,37 \quad \text{φυτά.} \end{aligned}$$

Ο μεγάλος αριθμός των φυτών, συγκρινόμενος με τον μικρό αριθμό των φύλλων ανά φυτό και τον πολύ μικρό αριθμό των προσδιορισμών ανά φύλλο, είναι φυσικά αποτέλεσμα της μεγάλης διακύμανσης μεταξύ των φυτών συγκριτικά με εκείνες των φύλλων και των προσδιορισμών.

**Περίπτωση Β.** Εάν επιθυμούμε η διακύμανση της μέσης τιμής να μην υπερβαίνει την τιμή  $s_{\bar{x}}^2 = {}_3V = 0,05$ , -αντί 0,1050- τότε η (1.48), δίδει:

$${}_3N_1 = \frac{(0,6043)(2,9956)}{(0,05)\sqrt{10}} = \frac{1,8102}{0,1581} = 11,45 \quad \text{φυτά.}$$

Με οδηγό τα παραπάνω αριθμητικά παραδείγματα και εφαρμόζοντας τους τύπους (1.49) έως (1.55), μπορούμε να υπολογίσουμε το μέγεθος του δείγματος σε κάθε επίπεδο δειγματοληψίας τεσσάρων επιπέδων. Με λίγη προσπάθεια, εάν προσέξουμε τη λογική της δομής των τύπων που δώσαμε για δύο, τρία και

τέσσερα επίπεδα, θα μπορούσαμε να τους επεκτείνουμε και σε περισσότερα επίπεδα ιεραρχικής δειγματοληψίας, που βέβαια είναι σπάνιο να αντιμετωπίσουμε τέτοιες περιπτώσεις στην πράξη.

## 6. Ιεραρχική Δειγματοληψία με Άνισο Αριθμό Επαναλήψεων

Υπάρχουν περιπτώσεις ιεραρχικής δειγματοληψίας υπαίθρου που για κάποιο λόγο έχουμε άνισο αριθμό επαναλήψεων (δειγμάτων) στα διάφορα ιεραρχικά επίπεδα. Ο λόγος μπορεί να είναι η σπανιότητα του υλικού, η καταστροφή κάποιων δειγμάτων, το απαγορευτικό κόστος κ.τ.λ. Στις περιπτώσεις αυτές ο υπολογισμός των στοιχείων του πίνακα της Ανάλυσης της Παραλλακτικότητας (ANOVA) παρουσιάζει κάποιες ιδιαιτερότητες (δυσκολίες), την αντιμετώπιση των οποίων απλά θα εκθέσουμε κατωτέρω με ένα αριθμητικό παράδειγμα χωρίς να υπεισέλθουμε στην αντίστοιχη θεωρητική εμβάθυνση.

Έστω ότι έχουμε τα δεδομένα μιας δειγματοληψίας σε τρεις περιοχές (Α, Β και Γ) στις οποίες εξετάσαμε 3, 3 και 2 δένδρα αντίστοιχα. Από κάθε δένδρο ελήφθησαν 1 ή 2 ή 3 δείγματα, όπως φαίνεται στον Πίνακα που ακολουθεί. Έτσι από τις περιοχές Α, Β, και Γ πάρθηκαν 6, 5 και 3 δείγματα αντίστοιχα και οι τιμές της μεταβλητής που εξετάσαμε είναι αυτές που δίδονται στο σώμα του Πίνακα:

Περιοχές	Α			Β			Γ		Σύνολο
Δένδρα	I	II	III	I	II	III	I	II	
Δείγματα	8	9	8	6	8	8		9	
	–	10	9	9	9	–		10	
	–	–	10	–	–	–	9	–	
Σύνολο/δένδρο	8	19	27	15	17	8	9	19	
Σύνολο/περιοχή	54			40			28		122

Από τα δεδομένα αυτά υπολογίζουμε τα απαιτούμενα SS (αθροίσματα τετραγώνων) του πίνακα της Ανάλυσης Παραλλακτικότητας ως εξής:

$$\begin{aligned}
 \text{SS μεταξύ περιοχών} &= \\
 &= \frac{54^2}{6} + \frac{40^2}{5} + \frac{28^2}{3} - \frac{122^2}{14} \\
 &= 486 + 320 + 261,3 - 1063,14 = 1067,3 - 1063,14 = 4,16
 \end{aligned}$$

με  $3 - 1 = 2$  βαθμούς ελευθερίας.

$$\begin{aligned}
 \text{SS μεταξύ δένδρων (αγνοώντας περιοχή)} &= \\
 &= \frac{8^2}{1} + \frac{19^2}{2} + \frac{27^2}{3} + \frac{15^2}{2} + \frac{17^2}{2} + \frac{8^2}{1} + \frac{9^2}{1} + \frac{19^2}{2} - \frac{122^2}{14} \\
 &= 1070 - 1063,14 = 6,86
 \end{aligned}$$

με  $8 - 1 = 7$  βαθμούς ελευθερίας.

SS μεταξύ δένδρων εντός περιοχών = SS δένδρων – SS περιοχών =  $6,86 - 4,16 = 2,7$

με  $7 - 2 = 5$  βαθμούς ελευθερίας.

SS συνολικό =  $(1)(6^2) + (4)(8^2) + (6)(9^2) + (3)(10^2) - \frac{122^2}{14} = 1078 - 1063,14 = 14,86$

με  $14 - 1 = 13$  βαθμούς ελευθερίας.

SS σφάλματος =  $14,86 - 6,86 = 8$

με  $13 - 7 = 6$  βαθμούς ελευθερίας.

Τα στοιχεία αυτά δίδονται στη γνωστή μορφή Πίνακα Παραλλακτικότητας (ANOVA):

Πηγή παραλλακτικότητας	Βαθμοί ελευθερίας	Άθροισμα τετραγώνων, SS	Μέσο τετράγωνο, MS	F
Μεταξύ περιοχών	$3 - 1 = 2$	4,16	2,08	2,14
Δένδρα στις περιοχές	$7 - 2 = 5$	2,7	0,54	
Δείγματα στα δένδρα (Σφάλμα)	$13 - 7 = 6$	8,0	1,33	
Σύνολο	13	14,86		

Βλέπουμε ότι στο παράδειγμα αυτό εμφανίζεται το «ανώμαλο» φαινόμενο το  $MS$  (0,54) των δένδρων να είναι μικρότερο του  $MS$  του «σφάλματος» (1,33). Έτσι δεν υπολογίζουμε  $F$  για δένδρα αλλά αθροίζουμε τα  $SS$  δένδρων και δειγμάτων (καθώς και τους αντίστοιχους βαθμούς ελευθερίας) και έχουμε 10,7 με 11 β.ε., δηλαδή  $MS = 10,7/11 = 0,97$ . Άρα η μεταξύ των περιοχών τιμή του  $F = 2,08/0,97 = 2,14$ . Από τους Πίνακες της  $F$  κατανομής βλέπουμε ότι η τιμή αυτή, για 2 και 11 β.ε., δεν είναι στατιστικά σημαντική (ακόμα και για επίπεδο σημαντικότητας 5 % η τιμή του  $F$  θα έπρεπε να ήταν 3,98 ή μεγαλύτερη). Επομένως οι τιμές της μεταβλητής που μελετούμε δεν διαφέρουν σημαντικά μεταξύ τους στις τρεις περιοχές.

## 7. Δειγματοληψία με Βάση την Έκταση

Στις μεθόδους και τεχνικές δειγματοληψίας που εκθέσαμε ανωτέρω, η στρατηγική ήταν να εξετάσουμε έναν αριθμό (δείγμα) διακριτών δειγματοληπτικών μονάδων (π.χ. καρπούς, δένδρα κ.τ.λ.) και να εκτιμήσουμε τον κατά μονάδα αριθμό των «γεγονότων» (π.χ. αριθμό προνυμφών τού υπό μελέτη εντόμου).

Στην παρούσα προσέγγιση, με βάση την έκταση, ενδιαφερόμαστε να εκτιμήσουμε τον συνολικό αριθμό  $N$  των ατόμων (π.χ. φυτών, δένδρων κ.τ.λ.) σε μια πεπερασμένη συνολική έκταση  $A$ , υποθέτοντας ότι η χωροδιάταξη των ατόμων στον συνεχή χώρο είναι τυχαία. Εξετάζουμε έναν αριθμό δειγματοληπτικών τεμαχίων, η αθροιστική έκταση των οποίων είναι ένα κλάσμα  $p$  της συνολικής έκτασης  $A$ . Άρα η έκταση της δειγματοληψίας μας είναι  $pA$ . Έστω ότι στην έκταση αυτή βρήκαμε  $n$  άτομα (π.χ. φυτά) του υπό μελέτη είδους.

Από τα παραπάνω είναι προφανές ότι μια εκτίμηση  $\hat{N}$  του  $N$  είναι:

$$\hat{N} = \frac{n}{p} \quad (1.56)$$

και 
$$V(\hat{N}) = \frac{V(n)}{p^2} \quad (1.57)$$

Επίσης η στατιστική κατανομή (Διωνυμική) της τυχαίας μεταβλητής  $n$  είναι:

$$b(N : n, p) = \binom{N}{n} p^n q^{N-n}, \quad \text{όπου } q = 1 - p \quad (1.58)$$

Άρα 
$$V(n) = Npq \quad (1.59)$$

από την (1.57) και (1.59) έχουμε:

$$V(\hat{N}) = \frac{Npq}{p^2} = \frac{Nq}{p} \quad (1.60)$$

$$\text{ή } S_{\hat{N}} = \sqrt{\frac{Nq}{p}} \quad (1.61)$$

Όπως έχουμε ήδη αναφέρει και σε άλλες περιπτώσεις, στην πράξη αντικαθιστούμε στους τελευταίους τύπους το  $N$  με την εκτίμησή του, δηλ.  $\hat{N}$  για να έχουμε μια πρώτη εκτίμηση της διακύμανσης  $V(\hat{N})$  ή του σφάλματος  $S_{\hat{N}}$ . Φυσικά, εάν η διακύμανση αυτή είναι μεγάλη και δεν μας ικανοποιεί, θα πρέπει να αυξήσουμε το ποσοστό  $p$  της έκτασης της δειγματοληψίας.

Ο συντελεστής παραλλακτικότητας, με τον οποίο συνήθως αξιολογούμε την αξιοπιστία της εκτίμησής μας, είναι κατά τα γνωστά:

$$CV(\hat{N}) = \frac{1}{N} \sqrt{\frac{Nq}{p}} = \sqrt{\frac{q}{Np}} \equiv c \quad (1.62)$$

Άρα 
$$c^2 = \frac{q}{Np} = \frac{1-p}{Np} \quad \text{ή } c^2 Np + p = 1$$

και λύνοντας ως προς  $p$  έχω:

$$p = \frac{1}{1 + Nc^2} \quad (1.63)$$

Η σχέση αυτή αποκαλύπτει ότι όσο μεγαλύτερος είναι ο αριθμός  $N$  των ατόμων, τόσο μικρότερο ποσοστό  $p$  απαιτείται για την εκτίμηση με δεδομένο βαθμό αξιοπιστίας (εκφρασμένο με τον συντελεστή παραλλακτικότητας  $c$ ). Η μείωση του  $p$  με την αύξηση του  $N$  είναι φυσικά καμπυλόγραμμη (1.63) με τα κοίλα προς τα άνω. Η ανωτέρω σχέση μάς επιτρέπει επίσης να αποφασίσουμε το ποσοστό  $p$  της όλης υπό μελέτη έκτασης που πρέπει να εξετάσουμε ούτως ώστε, έχοντας μια κατά προσέγγιση εκτίμηση του  $N$ , να εκτιμήσουμε το μέγεθος του πληθυσμού με τον προαποφασισθέντα  $c$  βαθμό αξιοπιστίας. Η πράξη έδειξε ότι σε μελέτες αυτής της φύσεως ένα ποσοστό 5-10% της έκτασης είναι συνήθως αρκετό.

#### Αριθμητικό παράδειγμα 1.14

Έστω ότι από μια έκταση  $A = 100$  στρεμμάτων εξετάστηκε το 5% ( $p = 0,05$ ), δηλαδή  $pA = 5$  στρέμματα (π.χ. 5 τεμάχια του ενός στρέμματος), στα οποία βρέθηκαν συνολικά  $n = 40$  άτομα. Επομένως σε όλη την έκταση των 100 στρεμμάτων εκτιμάται (1.56) ότι υπάρχουν:

$$\hat{N} = \frac{40}{0,05} = 800 \text{ άτομα}$$

Από την (1.60) έχουμε:  $V(\hat{N}) = \frac{(800)(0,95)}{0,05} = 15200$

άρα  $S_{\hat{N}} = 123$  άτομα.

Από την (1.62) έχουμε:  $CV(\hat{N}) = \sqrt{\frac{0,95}{(800)(0,05)}} = 0,15 = 0,15$  ή 15%.

Εάν αυτός ο συντελεστής παραλλακτικότητας δεν ικανοποιεί (είναι μεγάλος) και θέλουμε να τον μειώσουμε, π.χ. στο 0,10, θα πρέπει να αυξήσουμε το ποσοστό της έκτασης που θα εξετάσουμε –σύμφωνα με την (1.63)– από 5% ( $p = 0,05$ ) στο:

$$p = \frac{1}{1 + (800)(0,10)^2} = 0,11 \text{ δηλ. στο } 11\%$$

δηλαδή πρέπει η δειγματοληψία να γίνει σε 11 στρέμματα αντί σε 5.

Γενικά, κρατώντας στην (1.63) σταθερές τις δύο από τις τρεις μεταβλητές εκτιμούμε την τρίτη. Για εκτενέστερη ανάπτυξη του θέματος ο αναγνώστης μπορεί να ανατρέξει στον Seber (1973).

## 8. Εισαγωγή στην Έννοια της Γεωστατιστικής

Κλείνοντας το κεφάλαιο της δειγματοληψίας είναι ίσως σκόπιμο να αναφερθούμε με πολύ σύντομα στην έννοια της γεωστατιστικής, η οποία τα τελευταία ιδίως



χρόνια βρίσκει εφαρμογές και στην οικολογία (Jongman et al. 1987, Flatman et al. In Keith, Editor 1988, Isaak and Srivastava 1989, Liebhold , Rossi and Kemp 1993, Aubry and Debouzie 2000, Wright et al. 2002, Nestel, 2005).

Στη συνήθη δειγματοληψία πληθυσμών γνωρίζουμε ότι ένας βασικός κανόνας είναι ότι το δείγμα μας πρέπει να ληφθεί *τυχαία* για να μπορέσουμε να χρησιμοποιήσουμε τη στατιστική θεωρία στην ανάλυση των δεδομένων μας. Έτσι, εάν θέλουμε π.χ. να εκτιμήσουμε τη μέση πυκνότητα των φυτών μιας περιοχής, με τη γνωστή μέθοδο των «τετραγώνων», θα πρέπει οι θέσεις των τετραγώνων να είναι τυχαίες στον χώρο ούτως ώστε να μην υπάρχει το στοιχείο της υποκειμενικότητας και να μπορέσουμε να χρησιμοποιήσουμε τη στατιστική επεξεργασία των δεδομένων.

Μια άλλη προσέγγιση είναι η ονομαζόμενη *συστηματική δειγματοληψία*, κατά την οποία οι θέσεις δειγματοληψίας επιλέγονται από τον ερευνητή, η υποκειμενικότητα του οποίου –στην καλύτερη περίπτωση– δεν μπορεί να αποκλειστεί τελείως.

Μια τρίτη προσέγγιση είναι να πάρουμε δείγματα σε *τακτές αποστάσεις* πάνω σε διατομές (transects). Με αυτόν τον τρόπο οι θέσεις των δειγμάτων δεν είναι πλέον υποκειμενικές. Επίσης, ο τρόπος αυτός έχει το πλεονέκτημα ότι καλύπτουμε την υπό μελέτη έκταση (ιδίως εάν είναι μεγάλη) ομοιόμορφα και έτσι αποφεύγεται το ανεπιθύμητο ενδεχόμενο να μην πάρουμε δείγμα από κάποιο σχετικά μεγάλο τμήμα της έκτασης, το οποίο ενδεχομένως έχει κάποια ιδιαιτερότητα. Τέλος, τα δεδομένα μιας τέτοιας τακτών αποστάσεων δειγματοληψίας προσφέρονται για κατασκευή πληθυσμιακών χαρτών.

Η εκτίμηση της σταθερής απόκλισης της υπό μελέτη μεταβλητής (π.χ. πυκνότητα φυτών) από μια δειγματοληψία στον χώρο, προϋποθέτει ότι τα δείγματα είναι ανεξάρτητα μεταξύ τους. Η ιδιότητα αυτή δεν μπορούμε να δεχθούμε ότι ισχύει χωρίς να την ελέγξουμε. Συνήθως δεν ισχύει. Δηλαδή, δείγματα που έχουν μικρή απόσταση μεταξύ τους, εμφανίζουν κατά κανόνα μεγαλύτερο συντελεστή συσχέτισης σε σχέση με δείγματα που απέχουν αρκετά μεταξύ τους. Μετά από κάποια απόσταση, τα δείγματα είναι πρακτικά ανεξάρτητα και η αυτοσυσχέτισή τους σχεδόν μηδενίζεται. Αλλά ποια είναι αυτή η απόσταση;

Την απόσταση αυτή και γενικότερα τη μείωση της συσχέτισης με την απόσταση των δειγμάτων, μπορούμε να τις εκτιμήσουμε αναλύοντας τα δεδομένα της δειγματοληψίας με τα οποία κατασκευάζουμε *χωρικά συσχετογράμματα* χρησιμοποιώντας τις μεθόδους της γεωστατιστικής. Το συσχετογράμμα δίνει τη συσχέτιση  $\rho(h)$  μεταξύ δειγμάτων που χωρίζονται με απόσταση  $h$ . Υπολογίζεται με τον τύπο:

$$\rho(h) = \frac{\sum z_1 z_2 - N_h M_h^2}{N_h S_h^2}$$

όπου, τα  $z_1$  και  $z_2$  είναι (στην περίπτωση μας) οι αριθμοί των φυτών σε δύο δείγματα που έχουν μεταξύ τους απόσταση  $h$  στον χώρο. Η άθροιση γίνεται για όλα τα ζεύγη δειγμάτων που απέχουν απόσταση  $h$ . Το  $N_h$  είναι ο αριθμός των ζευγών των δειγμάτων που απέχουν μεταξύ τους απόσταση  $h$ . Τα  $M_h$  και  $S_h$  είναι αντίστοιχα η μέση τιμή και η σταθερή απόκλιση των δειγμάτων που απέχουν απόσταση  $h$ . Σημειώνουμε ότι κάθε δείγμα σταθμίζεται με βάση τον αριθμό των ζευγών δειγμάτων στα οποία συμμετέχει.

Ο ανωτέρω συντελεστής συσχέτισης μειώνεται συνήθως με την απόσταση  $h$  μεταξύ των δειγμάτων, καμπυλόγραμμα με τα κοίλα προς τα άνω. Σε κάποια απόσταση  $a$  η συσχέτιση μηδενίζεται. Η απόσταση αυτή είναι η ελάχιστη που πρέπει να έχουν τα δείγματα μιας έρευνας για να μπορούν να εφαρμοστούν οι συνήθεις στατιστικές μέθοδοι στα δεδομένα.

Το ανωτέρω συσχετόγραμμα είναι *όλων των κατευθύνσεων* (omnidirectional), δηλαδή βασίζεται στις αποστάσεις που παίρνουμε προς κάθε κατεύθυνση. Υπάρχουν και τα συσχετογράμματα *κατεύθυνσης* (directional correlograms). Μπορεί να συμβεί τα δείγματα να συσχετίζονται περισσότερο αναφορικά με μία κατεύθυνση (π.χ. ΒΑ-ΝΔ) παρά με άλλες. Εάν το συσχετόγραμμα εξαρτάται από την κατεύθυνση, τότε η χωροδιάταξη ονομάζεται *ανισότροπη* (anisotropic). Εάν δεν εντοπίστηκε ανισοτροπισμός, τότε μπορούμε να χρησιμοποιήσουμε συσχετόγραμμα *όλων των κατευθύνσεων*.

Να σημειώσουμε, τέλος, ότι η συνάρτηση του συσχετογράμματος χρησιμεύει μόνο εάν δεν υπάρχουν *τάσεις* της πυκνότητας στην περιοχή (δηλαδή η πυκνότητα των οργανισμών να έχει την τάση να αυξάνει ή να μειώνεται προς κάποια κατεύθυνση). Εάν υπάρχει τέτοια τάση, τότε θα πρέπει να χρησιμοποιηθεί το ονομαζόμενο non-ergotic συσχετόγραμμα.