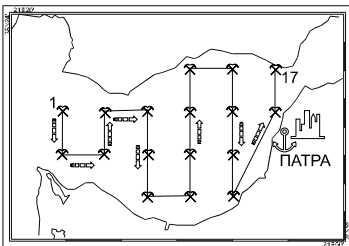


6.ΣΥΣΤΗΜΑΤΙΚΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ (SYSTEMATIC SAMPLING)

Η συστηματική δειγματοληψία είναι μια στρατηγική που μοιάζει απλή και λογική και επιλέγεται συχνά αυθόρμητα από μελετητές.. Η αυθόρμητη αυτή προτίμηση οφείλεται στο ότι διευκολύνει την επιλογή των δειγματοληπτικών μονάδων λόγω του απλού της πρωτοκόλλου. Για παράδειγμα αν επιθυμούμε να εκτιμήσουμε το μέσο βάρος των προβάτων μιας κτηνοτροφικής μονάδας ζυγίζοντας μόνο το ένα δέκατο (1/10) των ατόμων για πρακτικούς λόγους, τότε μπορούμε εύκολα κατά την είσοδο των ατόμων το απόγευμα να επιλέγουμε και να ζυγίζουμε ένα άτομο κάθε δέκα που περνούν από την είσοδο. Θα έχουμε έτσι ένα δείγμα το μέγεθος του οποίου είναι το 1/10 του πληθυσμού (που σ' αυτή την περίπτωση είναι το σύνολο των προβάτων της μονάδας). Συχνά επίσης εφαρμόζεται η συστηματική δειγματοληψία και σε μελέτες που καλύπτουν μεγάλες εκτάσεις τόσο στην ξηρά (ειδικά στις απ' αέρος παρατηρήσεις) όσο και στη θάλασσα. Σ' αυτή την περίπτωση η επιλογή της συστηματικής δειγματοληψίας γίνεται λόγω κυρίως της ευκολίας που παρέχει στα πληρώματα για την κάλυψη της υπό μελέτη έκτασης. Σ' αυτές τις περιπτώσεις ορίζεται εκ των προτέρων ένα σύνολο σημείων δειγματοληψίας (σταθμών) που απέχουν μεταξύ τους συγκεκριμένη απόσταση (π.χ. κάθε 20 km ή κάθε 10 ναυτικά μίλια στη θάλασσα). Οι σταθμοί αυτοί τοποθετούνται σε παράλληλες ευθύγραμμες διαδρομές που απέχουν μεταξύ τους δεδομένη και σταθερή απόσταση και διατρέχονται διαδοχικά. Παράδειγμα μιάς τέτοιας συστηματικής δειγματοληψίας φαίνεται στο σχήμα. **6.1.**



ΣΧΗΜΑ 6.1 Συστηματική δειγματοληψία στον Πατραϊκό κόλπο (17 σταθμοί).

Το μέγεθος του δείγματος που επιθυμούμε να μελετήσουμε καθορίζει και τον τρόπο επιλογής των ατόμων. Έτσι αν θέλουμε να σχηματίσουμε ένα δείγμα που να αντιπροσωπεύει το 1/p του

1	2	3								
4	5	6								
7	8	9								

ΣΧΗΜΑ 6.2 Επιλογή συστηματικού δείγματος:

- 1) Ο πληθυσμός (ή η έκταση που καλύπτει χωρίζεται σε 12 πρωτογενείς μονάδες ίδιου σχήματος και μεγέθους
- 2) Οι πρωτογενείς μονάδες διαιρούνται σε δευτερογενείς ίδιου αριθμού, σχήματος, μεγέθους και διάταξης
- 3) Από τις δευτερογενείς μονάδες μιας πρωτογενούς επιλέγεται τυχαία μια (στο παράδειγμα μας η μονάδα 3
- 4) Αυτή και η αντίστοιχη (με τις ίδιες συντεταγμένες) δευτερογενείς μονάδες όλων των πρωτογενών αποτελούν το συστηματικό δείγμα που αντιπροσωπεύει το 1/9 του πληθυσμού.

1	2	3								
4	5	6								
7	8	9								

Συστηματικό δείγμα ίδιου μεγέθους (1/9 του πληθυσμού) αλλά με διαφορετικό μέγεθος των πρωτογενών μονάδων (σχόλια στο κείμενο).

πληθυσμού, αρκεί από την λίστα των ατόμων του πληθυσμού ή από την σειρά προσέλευσης, ή εμφάνισης τους να επιλέγουμε ένα κάθε p άτομα ($p=N/n$ είναι το βήμα της δειγματοληψίας, N τα άτομα του πληθυσμού και n τα του δείγματος). Για την επιλογή δειγμάτων με συστηματικό τρόπο από πληθυσμούς που κατανέμονται στο χώρο και για τους οποίους δεν υπάρχει κατάλογος των ατόμων (σχεδόν όλοι οι φυσικοί πληθυσμοί) ακολουθούμε την εξής διαδικασία. Χωρίζουμε τον πληθυσμό μας (ή την έκταση που καλύπτει αυτός σε ισομεγέθεις και ίδιου σχήματος πρωτογενείς μονάδες. Στη συνέχεια κάθε πρωτογενής μονάδα χωρίζεται με το ίδιο τρόπο σε p δευτερογενείς. Από τις p μονάδες μίας πρωτογενούς μονάδος επιλέγεται τυχαία μια (χρήση υπολογιστή, πίνακας τυχαίων αριθμών κ.λ.π.) και στη συνέχεια σε κάθε πρωτογενή μονάδα επιλέγεται η δευτερογενής με τις ίδιες συντεταγμένες. Το σύνολο των δευτερογενών αυτών μονάδων αποτελεί ένα συστηματικό δείγμα ίσο με το $1/p$ του πληθυσμού στόχου.

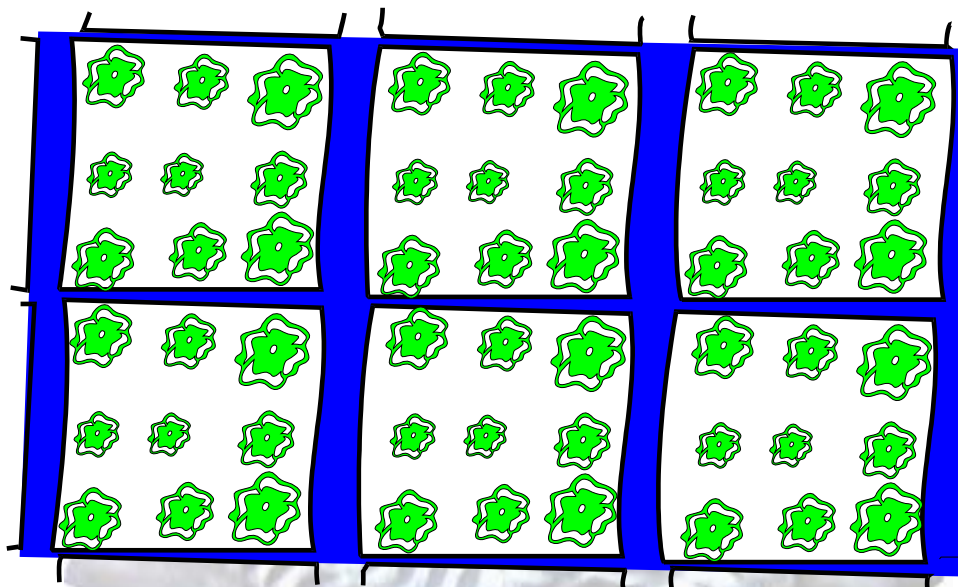
Είναι προφανές ότι το μέγεθος και κατά συνέπεια ο αριθμός των πρωτογενών μονάδων καθορίζει και τα χαρακτηριστικά του δείγματος. Μικρές και πολυπληθείς πρωτογενείς μονάδες οδηγούν σε μια καλύτερη “κάλυψη” του πληθυσμού. Καλύτερη κάλυψη σημαίνει ότι διατρέχουμε όλον τον πληθυσμό με λεπτομερή τρόπο και δεν αφήνουμε ακάλυπτες μεγάλες περιοχές. Σκεφτείτε στο σχήμα 6.2 να είχαμε μόνο δυο πρωτογενείς μονάδες. Το τελικό δείγμα πάλι θα ήταν ίσο με το $1/9$ του πληθυσμού όμως όπως φαίνεται και από το ίδιο σχήμα η πληροφορία θα προερχόταν μόνο από δυο περιοχές που θα αντιστοιχούσαν στις επιλεγμένες δευτερογενείς μονάδες. Σ’ αυτή την περίπτωση εάν ο πληθυσμός είχε μια έντονα συναθροιστική κατανομή (εναλλαγή περιοχών με υψηλές συγκεντρώσεις και κενών περιοχών) ο κίνδυνος ενός δείγματος με “ακραία” χαρακτηριστικά αυξάνει. Το να “διατρέχεται” καλά το δείγμα μειώνει τις πιθανότητες ακραίου δείγματος αλλά δυστυχώς αυξάνει συχνά το κόστος της δειγματοληψίας λόγω αυξημένων μετακινήσεων μεταξύ των δειγματοληπτικών μονάδων. Ένας “συμβιβασμός” των δυο αυτών κριτηρίων είναι στην πράξη απαραίτητος.

Η συστηματική δειγματοληψία έχει λοιπόν το πλεονέκτημα του απλού πρωτοκόλου και της ικανοποιητικής κάλυψης του πληθυσμού. Δυστυχώς όμως παρουσιάζει και σοβαρά μειονεκτήματα που αναλύονται στην ακόλουθη παράγραφο.

6.1. ΒΑΣΙΚΑ ΠΡΟΒΛΗΜΑΤΑ ΤΗΣ ΣΥΣΤΗΜΑΤΙΚΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Η συστηματική δειγματοληψία είναι εξαιρετικά μεροληπτική στην περίπτωση που ο πληθυσμός έχει περιοδικά χαρακτηριστικά. Φυσικά η μεγάλη μεροληψία είναι ανεπιθύμητη. Το βασικό αυτό πρόβλημα μπορεί να γίνει κατανοητό στα ακόλουθα παραδείγματα.

Στο σχήμα 6.3 παρουσιάζεται ένα μέρος από μια καλλιέργεια είδους φυτού. Η καλλιεργήσιμη περιοχή είναι χωρισμένη σε περιοχές που καθορίζονται από τα αρδευτικά αυλάκια που χωρίζουν την όλη περιοχή σε ισομεγέθη τεμάχια. Το συγκεκριμένο είδος φυτού είναι ιδιαίτερα απαιτητικό σε νερό και έτσι τα άτομα που βρίσκονται κοντά στα αυλάκια εμφανίζουν υψηλότερους ρυθμούς αύξησης και αυτο φαίνεται και στο σχήμα. Μια

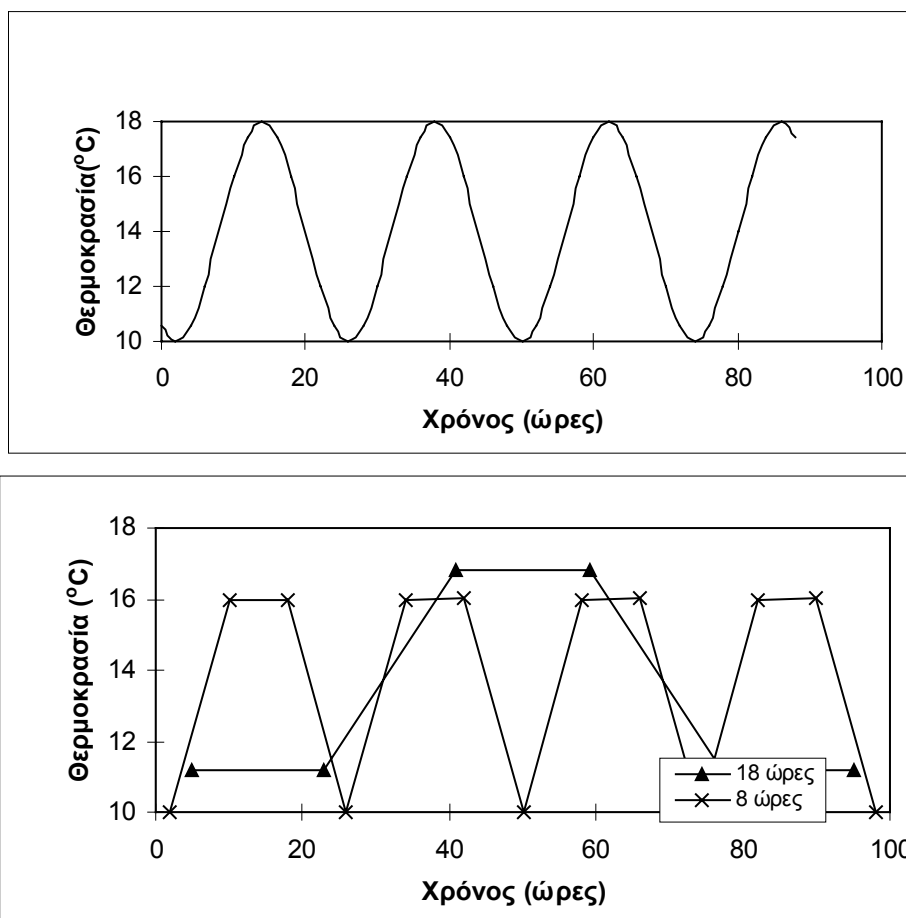


ΣΧΗΜΑ 6.3 Κατανομή φυτών σε μια καλλιέργεια. Είναι προφανές ότι το μέγεθος του φυτού εξαρτάται από την θέση του. Τα φυτά που βρίσκονται κοντά στα αρδευτικά αυλάκια έχουν μεγαλύτερο μέγεθος με τα άτομα που βρίσκονται στη συμβολή των αυλακιών να υπερτερούν σημαντικά. Ο περιοδικός αυτός χαρακτήρας εγκυμονεί κινδύνους μεγάλης μεροληψίας σε περίπτωση συστηματικής δειγματοληψίας.

συστηματική λοιπόν δειγματοληψία με πρωτογενείς μονάδες τα καλλιεργήσιμα τεμάχια και δευτερογενείς τα φυτά που έχουν την ίδια διάταξη σε κάθε τεμάχιο θα οδηγήσει σε υπερ-εκτίμηση ή υπο-εκτίμηση του μέσου μεγέθους των φυτών και φυσικά και της συνολικής ποσότητας.

Ο κίνδυνος εφαρμογής της συστηματικής δειγματοληψίας σε πληθυσμούς που εμφανίζουν περιοδικότητα γίνεται πιο κατανοητός στο παράδειγμα του σχήματος 6.4. Ας υποθέσουμε ότι επιθυμούμε να εκτιμήσουμε την μέση ημερήσια θερμοκρασία σε ένα συγκεκριμένο σημείο στο οποίο δεν μπορεί να τοποθετηθεί αυτόματος καταγραφέας (μια φωλιά για παράδειγμα). Η παράμετρος αυτή δείχνει έντονες ημερονύκτιες διακυμάνσεις με τη μέγιστη τιμή της τις πρώτες μεταμεσημβρινές ώρες και την ελάχιστη στο μέσο της νύχτας. Η πραγματική διακύμανση της θερμοκρασίας φαίνεται στο σχήμα 6.4 που δείχνει καθαρά τον 24ωρο περιοδικό χαρακτήρα της παραμέτρου. Το κάτω διάγραμμα του ίδιου σχήματος δείχνει τα αποτελέσματα που θα καταγράψουμε εάν η δειγματοληψία μας έχει περιοδικό χαρακτήρα με περίοδο 18

και 8 ωρών αντίστοιχα (το πρώτο βήμα των 18 ωρών μπορεί να οφείλεται στην δυσκολία πρόσβασης και μέτρησης ενώ το δευτερο είναι συνήθης συχνότητα που βασίζεται στις αλλαγές βάρδιας κάθε δωρο).



ΣΧΗΜΑ 6.4. Άνω: Χρονοσειρά θερμοκρασιών σε μια περιοχή που χαρακτηρίζεται από έντονες περιοδικές (ημερονύκτιες) διακυμάνσεις. Κάτω: Αποτελέσματα δειγματοληπτικών μετρήσεων στην ίδια περιοχή. Συστηματική δειγματοληψία με βήμα (χρόνο μεταξύ διαδοχικών μετρήσεων) 18 και 8 ώρες. Με πραγματική μέση θερμοκρασία 14°C οι εκτιμήσεις είναι 13,06 και 13,69 αντίστοιχα.

Η συστηματική αυτή δειγματοληψία όχι μόνο δεν βοηθά στην κατανόηση του βασικού χαρακτήρα του φαινομένου αλλά δίνει και λανθασμένες (μεροληπτικές εκτιμήσεις της μέσης τιμής (13,06 και 13,69 αντίστοιχα με πραγματική μέση τιμή 14°C). Ο κίνδυνος λάθους μεγαλώνει όσο η περίοδος της συστηματικής δειγματοληψίας πλησιάζει την περίοδο του φαινομένου. Έτσι στο ανωτέρω παράδειγμα αν οι μετρήσεις άρχιζαν το μεσημέρι της πρώτης μέρας και συνεχίζονταν κάθε 24 ώρες θα κατέγραφαν συνεχώς την ίδια τιμή που θα ήταν και η μέγιστη (18°C). Όλες οι τιμές του δείγματος θα ήταν ίδιες και φυσικά δεν θα έδιναν πληροφορία. Θα ήταν σα να είχαμε μια μόνο παρατήρηση. Και φυσικά όταν οι περίοδοι του φαινομένου και της δειγματοληψίας

είναι ίδιες, η τιμή της εκτίμησης εξαρτάται από την επιλογή της πρώτης μέτρησης (στη περίπτωση του παραδείγματος από την ώρα έναρξης των μετρήσεων. Αν άρχιζαν μετά τα μεσάνυχτα θα είχαμε την εντύπωση μιας σταθερής θερμοκρασίας 10°C.

Στη φύση τα φαινόμενα με περιοδικό χαρακτήρα είναι αρκετά συχνά (παλίρροια, ημερονύκτιες και εποχιακές αλλαγές κ.λ.π.) γι' αυτό και η επιλογή της συστηματικής δειγματοληψίας καθώς και του βήματος της πρέπει να γίνονται με μεγάλη προσοχή. Εάν τα αποτελέσματα μιας συστηματικής δειγματοληψίας δείχνουν ασυνήθιστα υψηλή ομοιογένεια τότε μια προσεκτική ανάλυση των χαρακτηριστικών του πληθυσμού σε σχέση με το βήμα της δειγματοληψίας επιβάλλεται.

Το δεύτερο πρόβλημα προέρχεται από το γεγονός ότι το συστηματικό δείγμα δεν είναι ένα τυχαίο δείγμα του πληθυσμού. Αυτό δυσχεραίνει την εκτίμηση βασικών παραμέτρων και κυρίως της διασποράς. Κατ' αρχήν από τον τρόπο επιλογής των ατόμων του δείγματος φαίνεται ότι πρόκειται για μια ειδική περίπτωση δισταδιακής δειγματοληψίας (πρωτογενείς μονάδες που περιέχουν δευτερογενείς από τις οποίες επιλέγεται μία).

Από θεωρητικής πλευράς η συστηματική δειγματοληψία παραβιάζει βασικές συνθήκες της τυχαίας επιλογής. Αυτό γίνεται κατανοητό από το ακόλουθο παράδειγμα. Από έναν υποθετικό πληθυσμό 18 ατόμων y_1, y_2, \dots, y_{18} , θέλουμε να επιλέξουμε ένα συστηματικό δείγμα ίσο με το \bar{O} του πληθυσμού. Σύμφωνα με τη διαδικασία επιλογής που αναφέρθηκε στην προηγούμενη παράγραφο όλα τα δυνατά δείγματα που μπορεί να δώσει ο πληθυσμός αυτός φαίνονται στον πίνακα 6.1. Η πρώτη παρατήρηση αφορά στο μέγεθος των δειγμάτων: άλλα περιέχουν 5 και άλλα 4 άτομα. Αυτό οφείλεται στο ότι το N/n που είναι ίσο με το p δεν είναι ακέραιος. Όμως στην πράξη οι συνέπειες αυτής της διαφοράς είναι μηδαμινές όταν το δείγμα είναι μεγάλο ($n > 50$). Το βασικό πρόβλημα είναι άλλο. Από τον πίνακα φαίνεται ότι ο πληθυσμός μπορεί να δώσει μόνο τα 4 συγκεκριμένα δείγματα. Η πιθανότητα τα άτομα y_1 και y_5 να βρεθούν στο ίδιο δείγμα είναι 0 ενώ η πιθανότητα να βρεθούν μαζί τα y_1 και y_2 είναι 0. Αντίθετα σε ένα τυχαίο δείγμα η πιθανότητα εμφάνισης οποιουδήποτε ζεύγους ατόμων είναι η ίδια. Από τη στιγμή που τα άτομα του δείγματος δεν είναι ανεξάρτητα το ένα από το άλλο δεν μπορούν και να μας πληροφορήσουν για την ετερογένεια του πληθυσμού. Κατά συνέπεια ένα μόνο συστηματικό δείγμα ενώ μας δίνει μια εκτίμηση της μέσης τιμής του πληθυσμού δυστυχώς δεν μπορεί να μας δώσει εκτίμηση της διασποράς της μέσης αυτής τιμής. Είναι λοιπόν άχρηστο ένα συστηματικό δείγμα παρά τα πρακτικά του πλεονεκτήματα; Σε κάποιες περιπτώσεις που θα αναφερθούν στην επόμενη παράγραφο η συστηματική δειγματοληψία μπορεί να εφαρμοσθεί.

ΠΙΝΑΚΑΣ 6.1. Συστηματικά δείγματα από πληθυσμό 18 ατόμων

Δείγμα	Άτομα
1	$y_1 \quad y_5 \quad y_9 \quad y_{13} \quad y_{17}$
2	$y_2 \quad y_6 \quad y_{10} \quad y_{14} \quad y_{18}$
3	$y_3 \quad y_7 \quad y_{11} \quad y_{15}$
4	$y_4 \quad y_8 \quad y_{12} \quad y_{16}$

6.2. ΕΚΤΙΜΗΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΗΣ ΣΥΣΤΗΜΑΤΙΚΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Στην περίπτωση που τα άτομα του πληθυσμού έχουν συγκεκριμένα χαρακτηριστικά το συστηματικό δείγμα μπορεί να μας δώσει εκτίμηση της διασποράς της μέσης τιμής. Θυμίζουμε ότι τα “άτομα” είναι οι δειγματοληπτικές μονάδες και ο πληθυσμός στον οποίο αναφερόμαστε είναι ο δειγματοληπτικός πληθυσμός.

Οι περιπτώσεις αυτές είναι όταν:

- τα άτομα κατανέμονται τυχαία μέσα στον πληθυσμό
- ο πληθυσμός χαρακτηρίζεται από γραμμικές τάσεις αλλαγής

Φυσικά στην εφαρμογή των αντίστοιχων εκτιμητών κάθε απομάκρυνση από αυτές τις συνθήκες έχει και σοβαρές επιπτώσεις στην αμεροληψία και την ακρίβειά τους.

6.2.1. Πληθυσμός με τυχαία κατανομή

Όταν υπάρχουν βάσιμα στοιχεία ώστε η κατανομή των ατόμων στο χώρο ή η σειρά προσέλευσης των ατόμων να θεωρείται τυχαία τότε παρακάμπτονται οι παραπάνω σχολιασμένες δυσκολίες και το συστηματικό δείγμα μπορεί να θεωρηθεί ένα απλό τυχαίο δείγμα και έτσι εφαρμόζονται οι εκτιμητές του κεφαλαίου 3.

ΕΚΤΙΜΗΤΕΣ

Δεδομένα

$$y_1, y_2, y_3, \dots, y_i, \dots, y_n$$

που στην πραγματικότητα είναι οι τιμές

$$y_i, y_{i+p}, y_{i+2p}, \dots, y_{i+(n-1)p}$$

όπου p είναι το βήμα της δειγματοληψίας και i το τυχαία

επιλεγμένο άτομο από τα πρώτα p του πληθυσμού $p = \frac{N}{n}$

Παράμετροι του δείγματος

$$f = \frac{n}{N} \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

Εκτίμηση της μέσης τιμής (\hat{Y}) του πληθυσμού

$$\hat{Y} = \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad v(\hat{Y}) = s_y^2 = \frac{s_y^2}{n} (1-f)$$

$$P\{\bar{y} - t_{\alpha/2} s_{\bar{y}} < \bar{Y} < \bar{y} + t_{\alpha/2} s_{\bar{y}}\} = 1 - \alpha$$

Εκτίμηση του συνόλου (\hat{Y}) του πληθυσμού

$$\hat{Y} = N\bar{y}_c \quad v(\hat{Y}) = N^2 s_y^2 c$$

$$P\{\hat{Y} - t_{\alpha/2} \sqrt{v(\hat{Y})} < Y < \hat{Y} + t_{\alpha/2} \sqrt{v(\hat{Y})}\} = 1 - \alpha$$

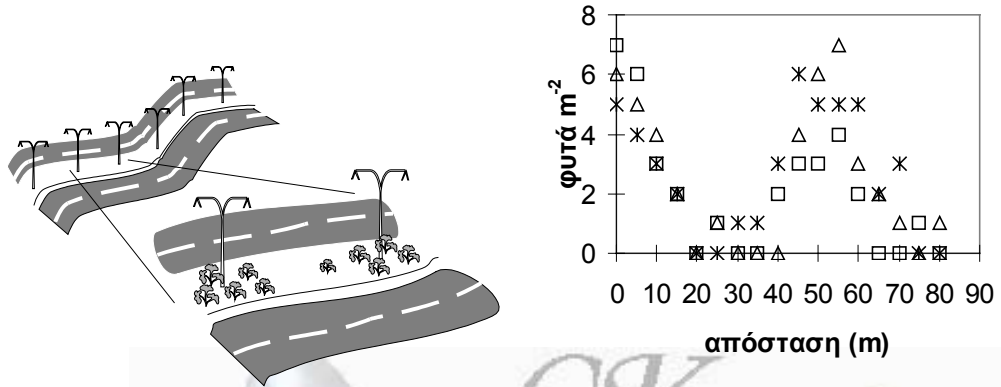
ΠΑΡΑΔΕΙΓΜΑ 6.1

Είδος φυτού αναπτύσσεται στη διαχωριστική νησίδα αυτοκινητοδρόμου. Για την ετήσια καταπολέμησή του απαιτείται εκτίμηση της συνολικής αφθονίας του ώστε να προβλεφθεί το κόστος της επιχείρησης. Προηγούμενες μελέτες έδειξαν ότι η συγκέντρωση ατόμων είναι σημαντική κοντά στις κολόνες φωτισμού. Οι κολόνες αυτές είναι τοποθετημένες κάθε 50 m καθ'όλο το μήκος της νησίδας που είναι 30 km. Το διάγραμμα του σχήματος 6.5 επιβεβαιώνει αυτήν την παρατήρηση. Η περιοχή που διατρέχει ο αυτοκινητόδρομος είναι αρκετά ομοιογενής τόσο από κλιματολογικής όσο και από τοπογραφικής πλευράς. Θεωρώντας τα στοιχεία αυτά και επίσης ότι το μέγεθος του δείγματος (μήκος νησίδας που θα αναλυθεί σε όλο της το πλάτος) δεν μπορεί να υπερβαίνει τα 1500 m, σχεδιάστε ένα πρωτόκολλο δειγματοληψίας.

Από τη στιγμή που δεν υπάρχει κανένα στοιχείο που να επιτρέπει τον ορισμό στρώσεων στον πληθυσμό στόχο μπορούμε να θεωρήσουμε χωρίς μεγάλο κίνδυνο λάθους ότι ο πληθυσμός έχει μια τυχαία κατανομή. Προσοχή, τυχαία θα είναι η κατανομή εάν το μέγεθος της δειγματοληπτικής μονάδας είναι ίσο η ακέραιο πολλαπλάσιο της απόστασης που χωρίζει τις δυο διαδοχικές κολόνες (50m). Εάν η δειγματοληπτική μονάδα έχει μέγεθος μικρότερο τότε η πυκνότητα του φυτού θα δείχνει έντονες διακυμάνσεις εξαρτώμενη από τη θέση (κοντά ή μακριά από κολόνα φωτισμού). Θα έχουμε δηλαδή τοιχαία ή συναθροιστική κατανομή ανάλογα με το μέγεθος της βασικής δειγματοληπτικής μονάδας.

Με βάση τα στοιχεία αυτά μια συστηματική δειγματοληψία με βασική μονάδα ένα τμήμα νησίδας μήκους 50m και με βήμα 20 θα επέτρεπε μια ικανοποιητική εκτίμηση σεβόμενη το ολικό κόστος της μελέτης. Στην πράξη λοιπόν επιλέγουμε έναν τυχαίο αριθμό από το 1 ως το 20 (έστω ότι είναι το 5), μετράμε τον αριθμό ατόμων από την κολόνα 5 ως την 6 (ή 25m πριν και μετά την κολόνα 5) και στη συνέχεια επαναλαμβάνουμε τη διαδικασία στις κολόνες 25, 45, 65 κ.ο.κ.. Έτσι θα αναλύσουμε 1500m συνολικό μήκος νησίδας με δειγματοληπτικές μονάδες που θα κατανέμονται ομοιόμορφα στον πληθυσμό. Πράγματι, 30km δρόμου με κολόνες κάθε 50m δίνουν ένα σύνολο 600 κολόνων.

Εξετάζοντας μια κάθε 20 (βημα=20) έχουμε ένα σύνολο 30 δειγματοληπτικών μονάδων που αντιπροσωπεύουν $30 \times 50\text{m} = 1500\text{m}$ νησίδας. Εκτός από την ομοιόμορφη κατανομή των δειγμάτων καθ'όλο το μήκος της διαδρομής, η συστηματική δειγματοληψία είναι και πιο εύκολη στην εφαρμογή της: ένα δείγμα κάθε 1km ή κάθε 20 κολόνες. Φυσικά όλα αυτά με την προϋπόθεση ότι οι αρχικές παραδοχές δεν απέχουν πολύ από την

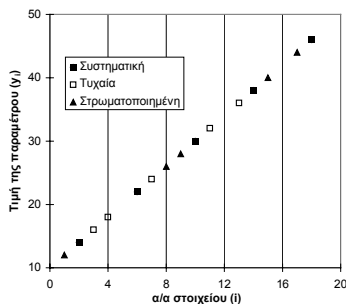


ΣΧΗΜΑ 6.5 Σχηματική παράσταση κατανομής είδους φυτού σε νησίδα αυτοκινητοδρόμου καθώς και αποτελέσματα μελετών που δείχνουν την πυκνότητα του φυτού σε σχέση με την απόσταση. Η επίδραση των κολονών φωτισμού που απέχουν μεταξύ τους 50 m είναι εμφανής αφού οι αποστάσεις μετρήθηκαν με σημείο αναφοράς μια τέτοια κολόνα.

πραγματικότητα.

6.2.1.1. Πληθυσμός με γραμμικές τάσεις αλλαγής

Όταν η τιμή μιας μεταβλητής (y_i) που χαρακτηρίζει τα άτομα ενός πληθυσμού εξαρτάται από τη θέση τους (i) και μάλιστα στην συγκεκριμένη περίπτωση είναι ανάλογη της θέσης τους ή της σειράς τους μέσα στον πληθυσμό ($y_i = \mu + \beta i$), τότε ένα μόνο συστηματικό δείγμα μπορεί να χρησιμοποιηθεί για την εκτίμηση παραμέτρων του πληθυσμού. Το δείγμα αυτό έχει το πλεονέκτημα να καλύπτει (διατρέχει) καλύτερα τον πληθυσμό από τις άλλες συνηθισμένες τεχνικές (σχήμα 6.6).



ΣΧΗΜΑ 6.6 Σύγκριση τριών τύπων δειγματοληψίας σε πληθυσμό που παρουσιάζει γραμμική τάση. Οι κάθετες γραμμές ορίζουν τις στρώσεις που έχουν το ίδιο μέγεθος με το βήμα της συστηματικής δειγματοληψίας ($p=4$).

ΕΚΤΙΜΗΤΕΣ

Δεδομένα

$$y_1, y_2, y_3, \dots, y_i, \dots, y_n$$

που στην πραγματικότητα είναι οι τιμές

$$y_i, y_{i+p}, y_{i+2p}, \dots, y_{i+(n-1)p}$$

όπου p είναι το βήμα της δειγματοληψίας και i το τυχαία

επιλεγμένο άτομο από τα πρώτα p του πληθυσμού $p = \frac{N}{n}$

Στον πληθυσμό αυτόν ισχύει $y_i = \mu + \beta \cdot i$ όπου β είναι η κλίση της γραμμικής σχέσης.

Παράμετροι του δείγματος

$$f = \frac{n}{N} \quad \bar{y} = \mu + \frac{np+1}{2}\beta$$

Εκτίμηση της μέσης τιμής (\hat{Y}) του πληθυσμού

$$\hat{Y} = \bar{y} = \mu + \frac{np+1}{2}\beta \quad v(\hat{Y}) = s_y^2 = \frac{(1-f)}{n} \sum_{u=1}^g \frac{d_u^2}{7.5g}$$

(εφαρμογή αυτής της σχέσης σε συστηματικές δειγματοληψίες φυσικών πληθυσμών έδωσε ικανοποιητικά αποτελέσματα αν και υπερ-εκτιμά ελαφρά την πραγματική διασπορά της μέσης τιμής)

Τα g και d_u υπολογίζονται ως ακολούθως

$g=n/9$ είναι ο αριθμός ομάδων των 9 ατόμων που μπορεί να δώσει το δείγμα

$$d_1=0.5y_1-y_2+y_3-y_4+y_5-y_6+y_7-y_8+0.5y_9$$

$$d_2=0.5y_9-y_{10}+y_{11}-y_{12}+y_{13}-y_{14}+y_{15}-y_{16}+0.5y_{17}$$

$$d_3=0.5y_{17}-y_{18}+ \dots +0.5y_{25}$$

$$d_u= \dots$$

$$d_g=0.5y_{n-8}-y_{n-7}+y_{n-6}-y_{n-5}+y_{n-4}-y_{n-3}+y_{n-2}-y_{n-1}+0.5y_n$$

$$P\{\bar{y} - t_{\alpha/2} s_y < \bar{Y} < \bar{y} + t_{\alpha/2} s_y\} = 1 - \alpha$$

Εκτίμηση του συνόλου (\hat{Y}) του πληθυσμού

$$\hat{Y} = N\bar{y}c \quad v(\hat{Y}) = N^2 s_y^2 c$$

$$P\{\hat{Y} - t_{\alpha/2} \sqrt{v(\hat{Y})} < Y < \hat{Y} + t_{\alpha/2} \sqrt{v(\hat{Y})}\} = 1 - \alpha$$

Αν ο μελετητής γνωρίζει αρκετά για την δομή του πληθυσμού τότε μπορεί να αναπτύξει ένα μαθηματικό μοντέλο που να περιγράφει τις εσωτερικές διακυμάνσεις του (το γραμμικό που αναφέρθηκε είναι το απλούστερο δυνατό). Με αυτό λοιπόν μπορεί να κατασκευάσει και μια συνάρτηση για την εκτίμηση της διασποράς της μέσης τιμής ή οποία θα είναι σχεδόν αμερόληπτη (unbiased) για το συγκεκριμένο μοντέλο διακύμανσης αλλά θα είναι πολύ μεροληπτική για πληθυσμούς με άλλη δομή. Όλα αυτά φυσικά ισχύουν εκτός από την περίπτωση των περιοδικών μεταβολών μέσα στον πληθυσμό που κάνουν τη συστηματική δειγματοληψία στην απλή της μορφή εντελώς αναποτελεσματική.

6.3. ΕΚΤΙΜΗΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΗΣ ΣΥΣΤΗΜΑΤΙΚΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Για να αποφύγουμε τις συνέπειες από μια λανθασμένη απόφαση για τον τύπο της εσωτερικής δομής του πληθυσμού, όπως αναφέρθηκε στην προηγούμενη παράγραφο, ή όταν η δομή του πληθυσμού είναι ιδιαίτερα πολύπλοκη μπορούμε να εφαρμόσουμε την επαναλαμβανόμενη συστηματική δειγματοληψία (repeated systematic sampling). Η αρχή είναι απλή. Αντί να πάρουμε ένα μόνο συστηματικό δείγμα μεγέθους n , επιλέγουμε m συστηματικά δείγματα μικρότερου μεγέθους που το άθροισμα των ατόμων τους (των δειγματοληπτικών μονάδων που περιέχουν δηλαδή) να είναι ίσο με n . Το μέγεθος των δειγμάτων αυτών είναι σταθερό και ίσο με $n_r = n/m$. Κάθε ένα από αυτά τα δείγματα δίνει μια εκτίμηση της μέσης τιμής και η διασπορά των μέσων αυτών τιμών είναι μια εκτίμηση της διασποράς του εκτιμητή της συστηματικής δειγματοληψίας. Έχουμε λοιπόν:

$$p = \frac{N}{n} \quad n_r = \frac{n}{m}$$

Από τα N/n_r πρώτα άτομα του πληθυσμού επιλέγω τυχαία m . Αυτά αποτελούν και τα αρχικά άτομα των m συστηματικών μου δειγμάτων. Τα υπόλοιπα άτομα των δειγμάτων αυτών επιλέγονται με βήμα ίσο με $m \cdot p$. Η διαδικασία αυτή φαίνεται στον πίνακα 6.2.

ΠΙΝΑΚΑΣ 6.2 Από ένα πληθυσμό 100 ατόμων θέλουμε συστηματικό δείγμα ίσο με το 20% του πληθυσμού. Άρα $n=0.2 \cdot N=20$ και $p=N/n=5$. Αντί για ένα θέλουμε να πάρουμε 5 συστηματικά δείγματα ($m=5$).

Έχουμε λοιπόν $n_r = n/m = 4$

Από τα N/n_r πρώτα άτομα επιλέγονται τυχαία m (δεύτερη κολόνα του πίνακα). Στη συνέχεια σε κάθε ένα από αυτά προσθέτοντας $m \cdot p$ βρίσκουμε τα υπόλοιπα άτομα του δείγματος

δείγμα				
1	1	26	51	76
2	7	32	57	82
3	12	37	62	87
4	21	46	71	96
5	23	48	73	99

Υπενθυμίζεται ότι οι αριθμοί αυτοί αντιπροσωπεύουν την θέση των ατόμων

ΕΚΤΙΜΗΤΕΣ

Δεδομένα

$$y_1, y_2, y_3, \dots, y_i, \dots, y_n$$

που στην πραγματικότητα είναι οι τιμές

$$y_i, y_{i+mp}, y_{i+2mp}, \dots, y_{i+(n-1)mp}$$

του ενός από τα m συστηματικά δείγματα

$$p = \frac{N}{n} \quad n_r = \frac{n}{m}$$

Παράμετροι του δείγματος

$$f = \frac{n}{N} \quad \bar{y}_j = \frac{\sum_{i=1}^{n_r} y_{ji}}{n_r} \quad \text{μέση τιμή καθενός από τα } m \text{ δείγματα}$$

Εκτίμηση της μέσης τιμής (\hat{Y}) του πληθυσμού

$$\hat{Y} = \bar{y} = \frac{\sum_{j=1}^m \bar{y}_j}{m} \quad v(\hat{Y}) = s_y^2 = \frac{\sum_{j=1}^m (\bar{y}_j - \bar{y})^2}{m(m-1)} (1-f)$$

$$P\{\bar{y} - t_{\alpha/2} s_y < \bar{Y} < \bar{y} + t_{\alpha/2} s_y\} = 1 - \alpha$$

με $m-1$ βαθμούς ελευθερίας

Εκτίμηση του συνόλου (\hat{Y}) του πληθυσμού

$$\hat{Y} = N\bar{y}c \quad v(\hat{Y}) = N^2 s_y^2 c$$

$$P\{\hat{Y} - t_{\alpha/2} \sqrt{v(\hat{Y})} < Y < \hat{Y} + t_{\alpha/2} \sqrt{v(\hat{Y})}\} = 1 - \alpha$$

6.4. ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΚΑΙ ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΤΗΣ ΣΥΣΤΗΜΑΤΙΚΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

↑ Πλεονεκτήματα ↑ C

- Η συστηματική δειγματοληψία είναι πολύ πιο εύκολη στην προτίμηση και την πραγματοποίησή της από τις άλλες στρατηγικές, ειδικά όταν πρόκειται για τη συλλογή πληροφορίας από πληθυσμούς που κατανέμονται σε μεγάλες εκτάσεις.
- Το πρωτόκολο συλλογής μειώνει στο ελάχιστο την πιθανότητα λάθους (π.χ. λάθος στίς συντεταγμένες σταθμού ή παράληψη παρατήρησης).
- Αποφεύγεται η μεροληπτική επιλογή ατόμων. Για παράδειγμα όταν ο ερευνητής προσπαθεί να εκτιμήσει το μέσο μέγεθος είδους ψαριού στην ιχθυόσκαλα έχει συχνά αυθόρμητα την τάση να επιλέγει για μέτρηση τα μεγαλύτερα άτομα.
- Όταν τα άτομα του πληθυσμού παρουσιάζονται με τυχαία σειρά ή κατανέμονται τυχαία στο χώρο τότε η συστηματική δειγματοληψία διατηρεί όλα τα πλεονεκτήματα της τυχαίας δειγματοληψίας (κεφάλαιο 3) αλλά επιπλέον είναι πιο ακριβής διότι οι παρατηρήσεις κατανέμονται πιο ομοιόμορφα και διατρέχουν (καλύπτουν)

↓ Μειονεκτήματα ↓

- Δεν επιτρέπει με σίγουρο τρόπο την εκτίμηση της διασποράς των εκτιμητών από τη συλλογή ενός μόνο συστηματικού δείγματος.
- Εάν ο πληθυσμός παρουσιάζει περιοδικές διακυμάνσεις τότε η στρατηγική αυτή εγκυμονεί μεγάλους κινδύνους και απαιτείται μεγάλη προσοχή στην επιλογή του βήματος δειγματοληψίας.
- Δεν επιτρέπει την επιλογή ατόμων με πιθανότητες που να σχετίζονται με την σχετική τους σπουδαιότητα και το ρόλο τους μέσα στον πληθυσμό.
- Το πλάνο αυτό δεν εκμεταλλεύεται εύκολα προϋπάρχουσες πληροφορίες για τη δομή και την κατανομή του πληθυσμού.
- Εκτός από την κατανομή παρατηρήσεων στο χώρο και τον χρόνο (σταθμούς) το πρωτόκολο αυτό απαιτεί γνώση της διάταξης των ατόμων στον πληθυσμό, πράγμα δύσκολο στις περισσότερες περιπτώσεις.

όλον τον πληθυσμό.

▪ Εάν ο πληθυσμός παρουσιάζει φαινόμενα αυτοσυσχέτισης (autocorrelation) τότε η συστηματική δειγματοληψία είναι πιο ακριβής διότι αποφεύγεται η πιθανότητα συλλογής δειγμάτων από μια συγκεκριμένη περιοχή και που φυσικά θα είναι επηρεασμένα από τα τοπικά χαρακτηριστικά. Τα φαινόμενα αυτοσυσχέτισης είναι πολύ κοινά στη φύση όπου οι κλίμακες των φαινομένων είναι μεγάλες. Για παράδειγμα δυο μετεωρολογικοί σταθμοί που βρίσκονται κοντά δίνουν πληροφορίες που μοιάζουν πολύ μεταξύ τους σε σχέση με δυο σταθμούς που απέχουν πολύ μεταξύ τους.

▪ Επίσης και όταν υπάρχει γραμμική διακύμανση στον πληθυσμό η συστηματική δειγματοληψία είναι πιο ακριβής σε σχέση με την απλή τυχαία για τον προαναφερθέντα λόγο (παρουσίαση στο σχήμα 6.6).

