
Επιλογή Δείγματος

Απόστολος Βανταράκης
Καθηγητής Υγιεινής
avanta@upatras.gr

Δειγματοληψία

- Να κατανοηθούν:
 - Γιατί κάνουμε δειγματοληψία
 - Ορισμοί δειγματοληψίας
 - Αντιπροσωπευτικότητα
 - Κύριοι μέθοδοι δειγματοληψίας
 - Λάθη στη δειγματοληψία

Γιατί κάνουμε δειγματοληψία;

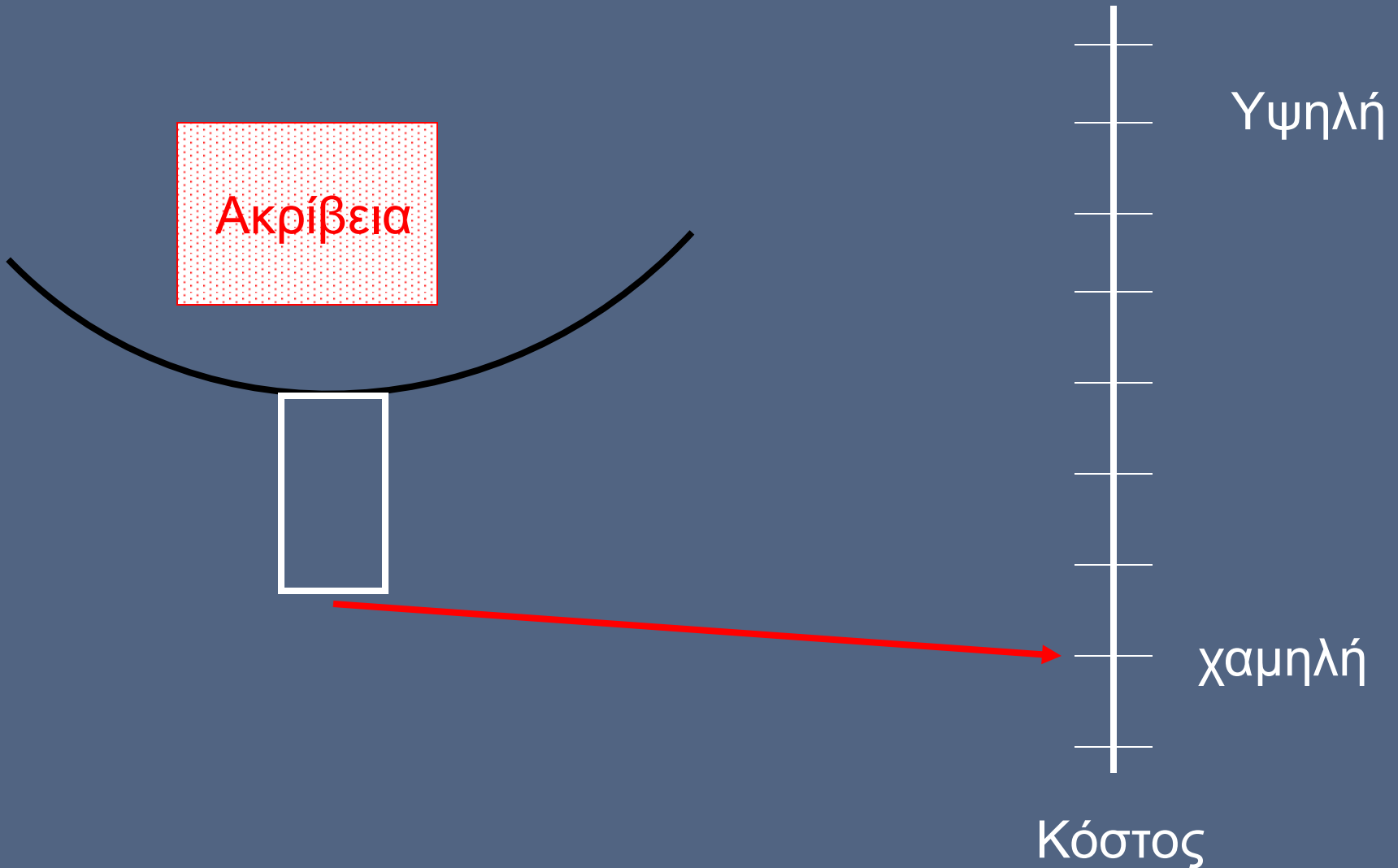
Να λάβουμε πληροφορίες από μεγάλους πληθυσμούς με:

- Μειωμένο κόστος
- Μειωμένο χρόνο πεδίου
- Αυξημένη ακρίβεια
- Ισχυρές μεθόδους

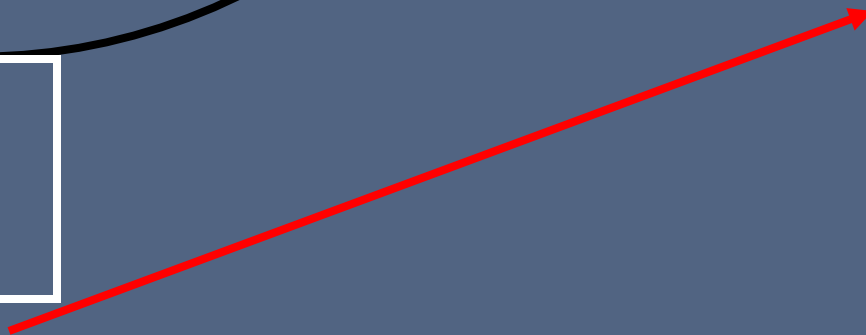
Ορισμός δειγματοληψίας

Διαδικασία με την οποία μερικά άτομα ενός δεδομένου πληθυσμού επιλέγονται σαν αντιπροσωπευτικά ολόκληρου του πληθυσμού

Δειγματοληψία



Δειγματοληψία

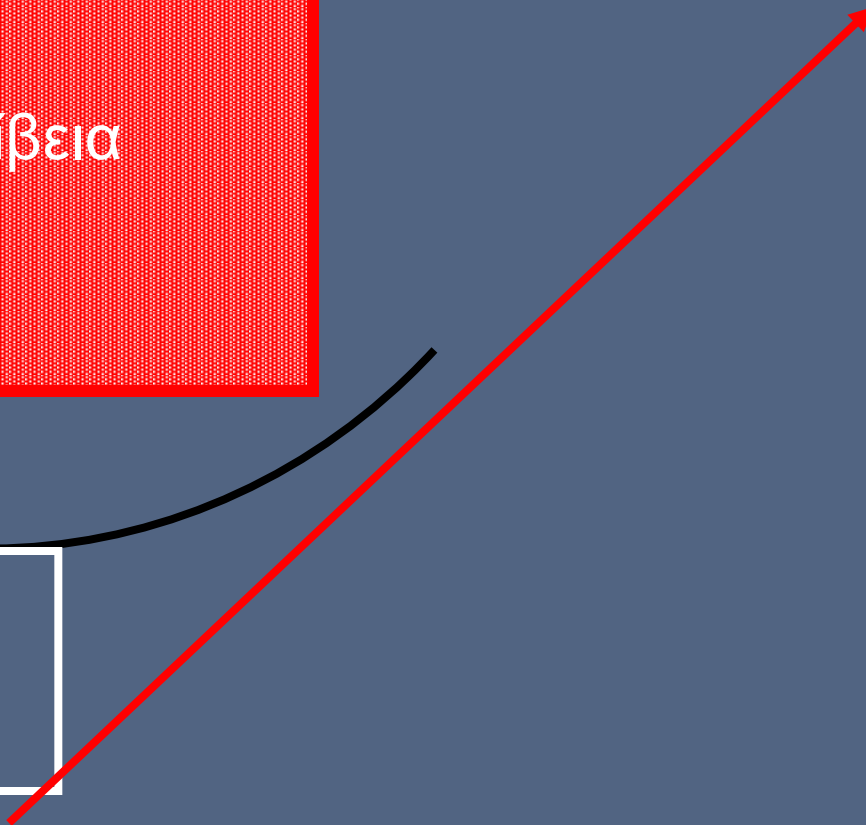


Υψηλή

χαμηλή

Κόστος

Δειγματοληψία



Υψηλή

Χαμηλή

Κόστος

Ορισμοί στην επιλογή δείγματος

Μονάδα δειγματοληψίας

- Αντικείμενο υπό παρατήρηση από το οποίο συλλέγονται οι πληροφορίες
 - παράδειγμα: παιδιά <5 ετών, εργαζόμενοι σε νοσοκομείο, περιστατικά

Κλάσμα δειγματοληψίας

- Ποσοστό του μεγέθους του δείγματος επί του πληθυσμού
 - παράδειγμα: 100 από 2000 (5%)

Ορισμοί στη δειγματοληψία

Δειγματοληπτικό πλαίσιο

- Κατάλογος όλων των μονάδων δειγματοληψίας από τις οποίες προκύπτει
 - κατάλογοι: π.χ. παιδιά < 5 ετών, νοικοκυρές

Δειγματοληπτικό σχήμα

- Μέθοδος επιλογής μονάδων δειγματοληψίας από το δειγματοληπτικό πλαίσιο
 - τυχαιοποιημένα, βολικό δείγμα...

Λάθη στη διαδικασία

- **Συστηματικό λάθος (or bias)**
 - Αντιπροσωπευτικότητα (εγκυρότητα)
 - Πληροφορίες για προκατάληψη (bias)
- **Λάθη δειγματοληψίας (τυχαίο λάθος)**
 - Ακρίβεια

Δειγματοληψία και Αντιπροσωπευτικότητα



Πληθυσμός στόχος → Πληθυσμός δειγματοληψίας → Δείγμα

Αντιπροσωπευτικότητα (εγκυρότητα)

- Το δείγμα θα πρέπει με ακρίβεια να αντανakλά την κατανομή της σχετικής ποικιλότητας στον πληθυσμό
 - άτομο (ηλικία, φύλο)
 - Τόπος π.χ. Αστικός-αγροτικός
 - Χρόνος π.χ εποχικότητα (Χειμώνα-Καλοκαίρι)
- Αντιπροσωπευτικότητα απαραίτητη για τη γενίκευση των αποτελεσμάτων
- Επιβεβαίωση της αντιπροσωπευτικότητας πριν την έναρξη
 - Επιβεβαίωση μετά την ολοκλήρωση

Προκατάληψη πληροφορίας (Information bias)

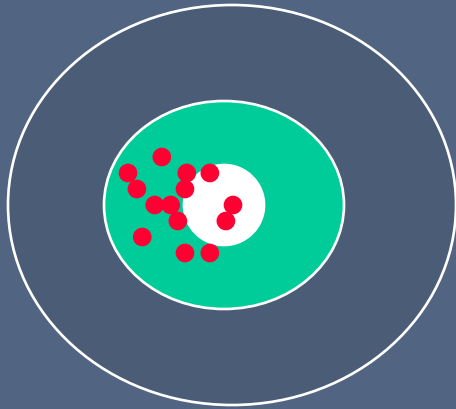
- Συστηματικό πρόβλημα στη συλλογή της πληροφορίας
 - Μη ακριβείς μετρήσεις
 - βαθμίδες, υπέρηχοι, εργαστηριακά tests...
 - Κακώς τοποθετημένες ερωτήσεις
 - φιλόδοξες, δεν προσφέρουν τις σωστές επιλογές απαντήσεων...

Λάθος δειγματοληψίας (τυχαιοποιημένο λάθος)

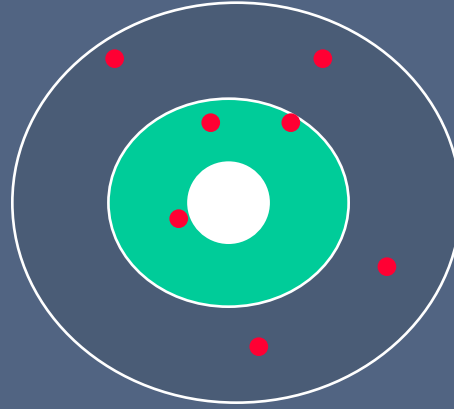
- Κανένα δείγμα δεν αποτελεί ακριβές είδωλο-καθρέφτη του πληθυσμού
 - Τυχαία διαφορά μεταξύ δείγματος και πληθυσμού από τον οποίο προκύπτει το δείγμα
 - Ακρίβεια
- Τυπικό λάθος που εξαρτάται από:
 - Μέγεθος δείγματος
 - Κατανομή του χαρακτήρα που μας ενδιαφέρει στον πληθυσμό
- Μέγεθος σφάλματος
 - Μπορεί να μετρηθεί σε πιθανά δείγματα
 - Τυπικό λάθος

Ποιότητα υπολογισμού επιλογής δείγματος

Ακρίβεια & εγκυρότητα

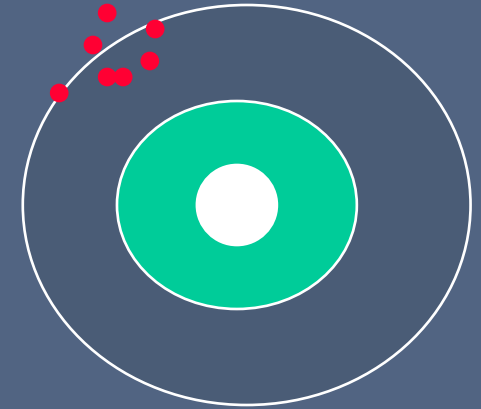


Καμία ακρίβεια



Τυχαίο
λάθος

Ακρίβεια αλλά
όχι εγκυρότητα



Συστηματικό
λάθος (bias)

Λάθη στην έρευνα: παράδειγμα

Μετρώντας το ύψος:

- Μετρώντας την ταινία που κρατείται διαφορετικά από τους διαφορετικούς ερευνητές
 - απώλεια ακρίβειας
 - Μεγάλο τυπικό λάθος
- Η ταινία πολύ κοντή
 - συστηματικό λάθος
 - Προκατάληψη Bias (δεν μπορεί να διορθωθεί μετά)

| |
|-----|
| |
| 179 |
| 178 |
| 177 |
| 176 |
| 175 |
| 174 |
| 173 |
| |

Τύποι δειγματοληψίας

- **Μη πιθανολογικά (Non-probability) δείγματα**

- Βολικά δείγματα

- Προκατειλημμένα (biased)
- Καλύτερο ή χειρότερο σενάριο (best or worst scenario)

- Υποκειμενικά δείγματα

- Βασίζονται στη γνώση
- Περιορισμοί σε χρόνο/χρήματα

- **Πιθανολογικά (Probability) δείγματα**

- Τυχαία

- Μόνο μέθοδοι που επιτρέπουν έγκυρα συμπεράσματα για τον πληθυσμό και μετρήσεις για το δειγματοληπτικό λάθος

Μη πιθανολογικά δείγματα

Βολικά δείγματα (ευκολία πρόσβασης)

- Δειγματοληψία χιονόμπαλας (φίλος φίλου...κλπ)
- Αιτιολογημένη δειγματοληψία (judgemental)
 - Διαλέγεις όποιον νομίζεις ότι πρέπει να είναι στη μελέτη

Πιθανότητα να επιλεγεί είναι άγνωστη
Φτηνότερο- αλλά ανίκανο να γενικευτεί, πιθανό για bias

Δείγματα πιθανοτήτων

- Τυχαία δειγματοληψία
 - Κάθε άτομο έχει μια γνωστή πιθανότητα να επιλεγεί
- Επιτρέπει την εφαρμογή της θεωρίας του στατιστικού δείγματος καταλήγει σε:
 - Γενίκευση
 - Έλεγχο υπόθεσης

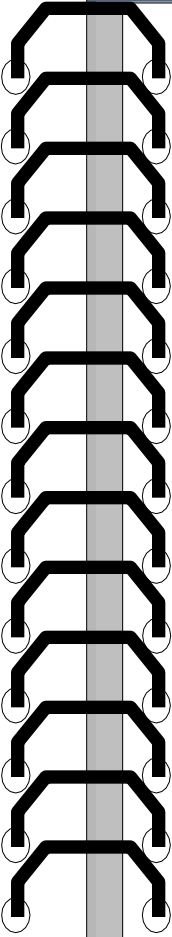
Μέθοδοι που χρησιμοποιούνται σε δείγματα πιθανοτήτων (τυχαία δειγματοληψία)

- Απλή τυχαία δειγματοληψία
- Συστηματική δειγματοληψία
- Στρωματοποιημένη δειγματοληψία
- Δειγματοληψία πολλών σταδίων
- Δειγματοληψία συστάδων

Απλή τυχαία δειγματοληψία

- Αρχή
 - Ίση ευκαιρία/πιθανότητα να επιλεγθεί κάθε μονάδα
- Διαδικασία
 - Παίρνουμε τον πληθυσμό στόχο
 - Κατάλογος όλων των δειγματοληπτικών μονάδων (“δειγματοληπτικό πλαίσιο”)
 - Αρίθμηση των μονάδων
 - Τυχαία επιλογή μονάδων

Απλή τυχαία επιλογή δείγματος



The diagram shows a vertical spinner with 48 numbered slots, each with a pointer. The spinner is divided into two halves by a vertical line. The numbers 1 through 48 are arranged in two columns, with 1-24 on the left and 25-48 on the right. The spinner is shown in a state where the pointer is pointing to the number 27.

| | | | |
|----|-----------------|----|----------------|
| 1 | Albert D. | 25 | Monique Q. |
| 2 | Richard D. | 26 | Réçine D. |
| 3 | Belle H. | 27 | Lucille L. |
| 4 | Raymond L. | 28 | Jérémy W. |
| 5 | Stéphane P. | 29 | Gilles D. |
| 6 | Albert T. | 30 | Renaud S. |
| 7 | Jean William V. | 31 | Pierre K. |
| 8 | André D. | 32 | Etienne M. |
| 9 | Jeremy W. | 33 | Marie M. |
| 10 | Anthony Q. | 34 | Gaétan Z. |
| 11 | James B. | 35 | Fidèle D. |
| 12 | Denis G. | 36 | Maria P. |
| 13 | Amanda L. | 37 | Anne Marie G. |
| 14 | Jennifer L. | 38 | Michel K. |
| 15 | Philippe K. | 39 | Gaston C. |
| 16 | Eve F | 40 | Alain M. |
| 17 | Priscilla O. | 41 | Olivier P. |
| 18 | Robert D | 42 | Geneviève M. |
| 19 | Brian F. | 43 | Berthe D. |
| 20 | Hellène H. | 44 | Jean Pierre P. |
| 21 | Isabelle R. | 45 | Jacques B. |
| 22 | Jean T. | 46 | François P. |
| 23 | Samanta D. | 47 | Dominique M. |
| 24 | Berthe L. | 48 | Antoine C. |

Απλή τυχαία επιλογή δείγματος

- Πλεονεκτήματα
 - Απλή
 - Το λάθος στην επιλογή δείγματος μετριέται εύκολα

- Πλεονεκτήματα
 - Χρειάζεται πλήρη λίστα των μονάδων
 - Δεν πετυχαίνει πάντα την καλύτερη αντιπροσωπευτικότητα
 - Οι μονάδες μπορούν να είναι διάσπαρτες και να μην είναι ιδιαίτερα προσβάσιμες
 - Ετερογενής πληθυσμός (μειονότητες)

Απλή τυχαία δειγματοληψία

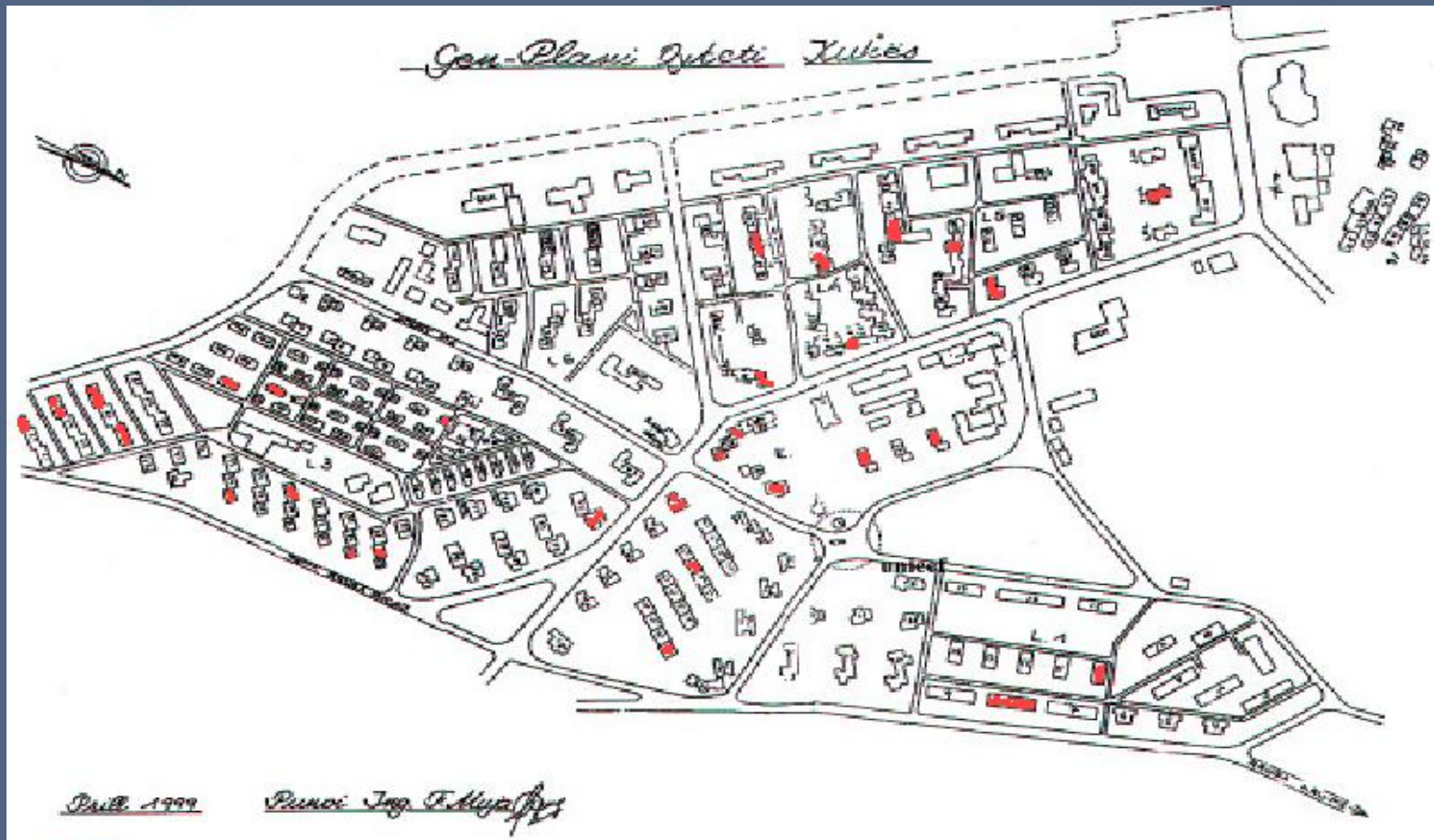
Παράδειγμα: αξιολόγηση του επιπολασμού τερηδόνας σε 1200 παιδιά που πηγαίνουν σχολείο

- Κατάλογος παιδιών που πηγαίνουν σχολείο
- Απαρίθμηση των παιδιών από το 1 έως το 1200
- Μέγεθος δείγματος = 100 παιδιά
- Τυχαία δειγματοληψία 100 αριθμών από το 1 έως το 1200

Πως επιλέγουμε τυχαία?

Όταν η συστηματική δειγματοληψία θα ήταν πιο αντιπροσωπευτική από ότι η τυχαία

Δειγματοληψία σπιτιών στην πόλη Kukes, 04/1999



Συστηματική δειγματοληψία

- Αρχή
 - Επιλογή μονάδων δειγματοληψίας σε τακτά διαστήματα
- Διαδικασία
 - Βάζουμε τις μονάδες σε σειρά
 - Διαιρούμε το συνολικό πληθυσμό από το σχεδιασμένο μέγεθος δείγματος (π.χ. $1200/60=20$)
 - Διαλέγουμε ένα τυχαίο σημείο έναρξης (για 20, το σημείο έναρξης θα ήταν ένας τυχαίος αριθμός μεταξύ 1 και 20)
 - Επιλέγετε μονάδες σε τακτικά διαστήματα (σε αυτή την περίπτωση κάθε περίπου 20 μονάδες)

Συστηματική Δειγματοληψία

- Πλεονεκτήματα

- Επιβεβαιώνει την αντιπροσωπευτικότητα στη λίστα
- Εύκολο να εφαρμοστεί

- Μειονεκτήματα

- Χρειάζεται συμπληρωμένη λίστα μονάδων
- Περιοδικότητα-υποκείμενο-πρότυπο μπορεί να είναι πρόβλημα (τα χαρακτηριστικά να συμβαίνουν σε τακτικά διαστήματα)

Συστηματική δειγματοληψία

| | | | | |
|----|-----------------|--|----|----------------|
| 1 | Albert D. | | 25 | Monique Q. |
| 2 | Richard D. | | 26 | Régine D. |
| 3 | Belle H. | | 27 | Lucille L. |
| 4 | Raymond L. | | 28 | Jérémy W. |
| 5 | Stéphane B. | | 29 | Gilles D. |
| 6 | Albert T. | | 30 | Renaud S. |
| 7 | Jean William V. | | 31 | Pierre K. |
| 8 | André D. | | 32 | Etienne M. |
| 9 | Jeremy W. | | 33 | Marie M. |
| 10 | Anthony Q. | | 34 | Gaétan Z. |
| 11 | James B. | | 35 | Fidèle D. |
| 12 | Denis G. | | 36 | Maria P. |
| 13 | Amanda L. | | 37 | Anne-Marie G. |
| 14 | Jennifer L. | | 38 | Michel K. |
| 15 | Philippe K. | | 39 | Gaston C. |
| 16 | Eve F. | | 40 | Alain M. |
| 17 | Priscilla O. | | 41 | Olivier P. |
| 18 | Robert D. | | 42 | Geneviève M. |
| 19 | Brian F. | | 43 | Berthe D. |
| 20 | Hellène H. | | 44 | Jean Pierre P. |
| 21 | Isabelle R. | | 45 | Jacques B. |
| 22 | Jean T. | | 46 | François P. |
| 23 | Samanta D. | | 47 | Dominique M. |
| 24 | Berthe L. | | 48 | Antoine C. |

Πιο πολύπλοκα συστήματα
δειγματοληψίας

Στρωματοποιημένη δειγματοληψία

- Πότε τη χρησιμοποιούμε
 - Πληθυσμός με διακριτές ομάδες
- Διαδικασία
 - Διαιρούμε (στρωματοποιούμε) το δειγματοληπτικό σχέδιο σε ομογενοποιημένες υποομάδες-στρώματα (strata) π.χ.. Ηλικιακά, αστικά/αγροτικά, περιοχές, εργασίες κλπ
 - Σχεδιάζουμε τυχαίο δείγμα σε κάθε στρώμα
 - Αν το μέγεθος των πληθυσμών είναι άνισο: η αναλογία των υποκειμένων από κάθε στρώμα (το ίδιο τμήμα δείγματος χρησιμοποιείται έτσι η πιθανότητα είναι αναλογική του μεγέθους του πληθυσμού)

Στρωματοποιημένη Δειγματοληψία

Επιλογή δείγματος με **πιθανότητα ανάλογη του μεγέθους**

| Περιοχή | Μέγεθος | Αναλογία | Μέγεθος δείγματος | Τμήμα Δείγματος |
|----------|---------|----------|-------------------------|-----------------|
| Αστική | 7000 | 70% | $1000 \times 0.7 = 700$ | 10 % |
| Αγροτική | 3000 | 30% | $1000 \times 0.3 = 300$ | 10 % |
| Σύνολο | 10000 | | 1000 | |

Στρωματοποιημένη δειγματοληψία

- Πλεονεκτήματα

- Μπορούν να αποκτηθούν πληροφορίες για όλο το πληθυσμό και για ξεχωριστά στρώματα
- Η ακρίβεια αυξάνεται αν η ποικιλότητα μέσα στα στρώματα είναι μικρότερη (ομογενοποιημένα) από ότι μεταξύ στρωμάτων

- Μειονεκτήματα

- Το λάθος στη δειγματοληψία είναι δύσκολο να μετρηθεί
- Διαφορετικά στρώματα είναι δύσκολο να ταυτοποιηθούν
- Απώλεια ακρίβειας αν οι μικροί αριθμοί σε ξεχωριστά στρώματα (λύνεται από δειγματοληψία ανάλογη με τον πληθυσμό του στρώματος)

Δειγματοληψία πολλαπλών σταδίων

Αρχές:

- Συνεχής δειγματοληψία
- Παράδειγμα :

δειγματοληπτική μονάδα = κατοικία

- 1st στάδιο: κλήρωση σε γειτονιές
- 2nd στάδιο: κλήρωση σε κτίρια
- 3rd στάδιο: κλήρωση σε σπίτια

Δειγματοληψία πολλών σταδίων

Επιλέγοντας τις μονάδες του πρώτου σταδίου με **πιθανότητα ανάλογη του μεγέθους**

Σενάριο

- 7 χωριά σε μια περιοχή
 - 3 χωριά να γίνει δειγματοληψία (1 βήμα μονάδες)
 - Σύνολο: 6000 άτομα
 - Λίστα αθροιστικής συχνότητας για όλα τα άτομα, τυχαίοι 3 αριθμοί
 - **μεταξύ 1 και 6000**
- e.g. 985, 3830, 4457

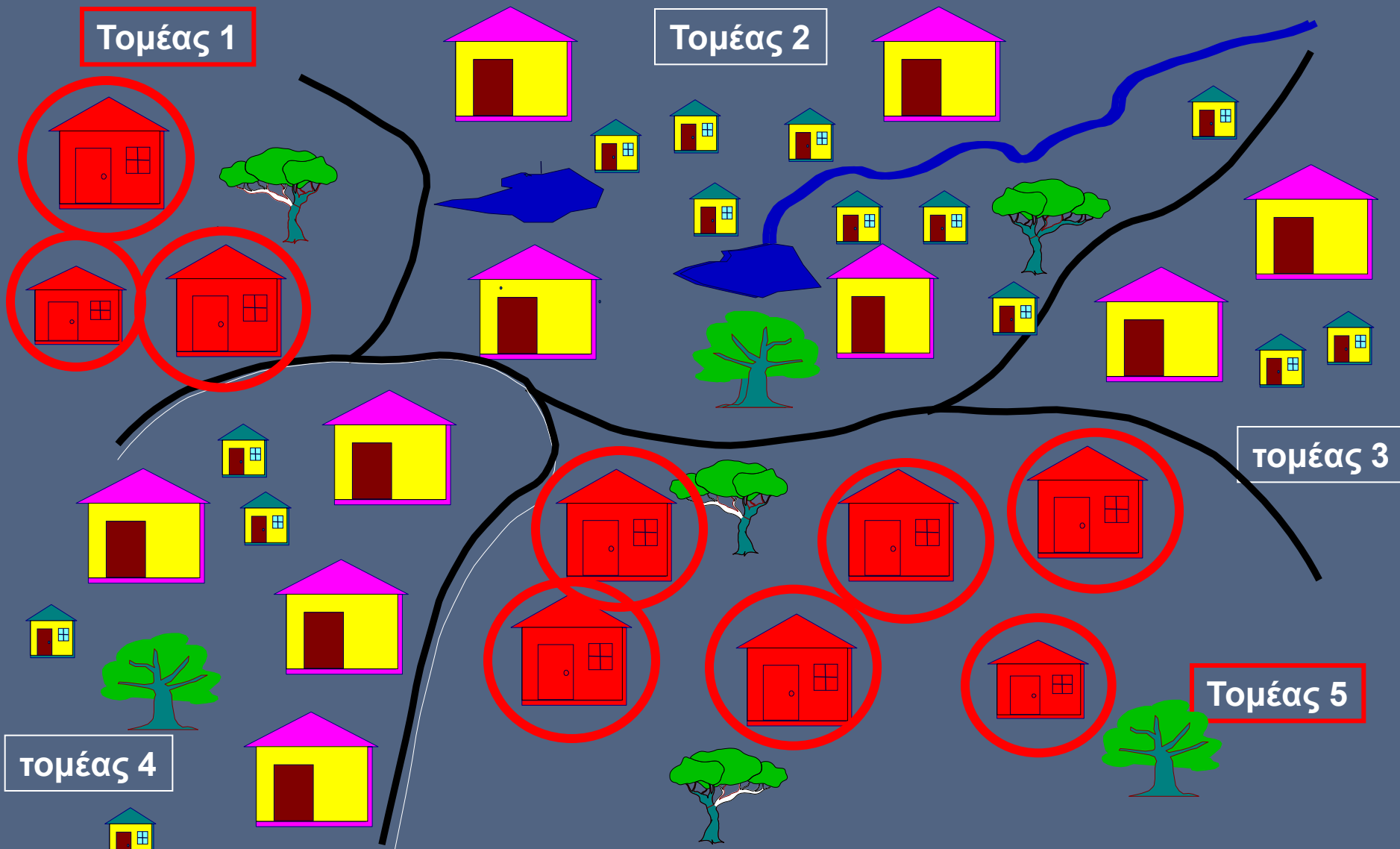
| Χωριό | Συχνότητα ατόμων | Αθροιστική συχνότητα |
|----------|------------------|----------------------|
| 1 | 30 | 30 |
| 2 | 400 | 430 |
| 3 | 1100 | 1530 |
| 4 | 500 | 2030 |
| 5 | 2000 | 4030 |
| 6 | 100 | 4130 |
| 7 | 1600 | 6000 |

Δειγματοληψία συστάδων (cluster)

- **Αρχή**

- Όλος ο πληθυσμός χωρίζεται σε ομάδες π.χ γειτονιές
- Μια δειγματοληψία πολλών σταδίων όπου όλες οι μονάδες στο χαμηλότερο επίπεδο περιλαμβάνονται στο δείγμα
- Τυχαίο δείγμα λαμβάνεται από αυτές τις ομάδες (“clusters”)
- Μέσα σε επιλεγμένες ομάδες, όλες οι ομάδες π.χ. περιλαμβάνονται τα σπίτια (ή τυχαίο δείγμα αυτών των μονάδων)
- Παρέχει λογιστικό πλεονέκτημα

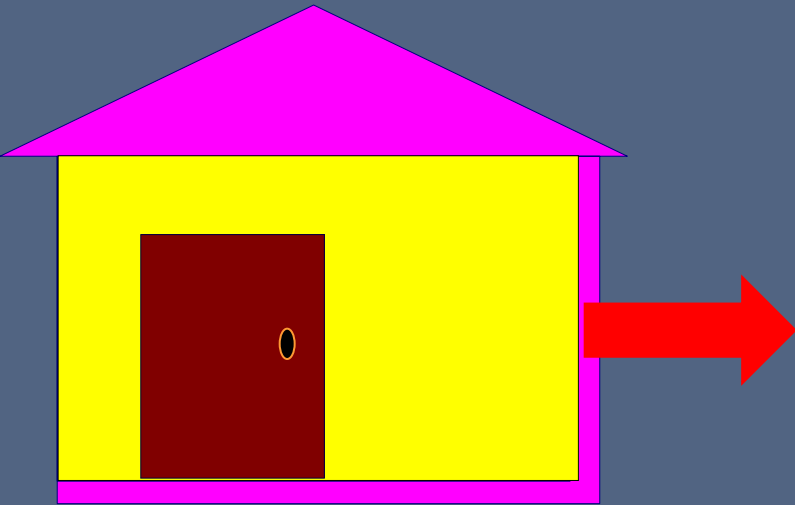
Παράδειγμα: δειγματοληψία σε 2-επίπεδα



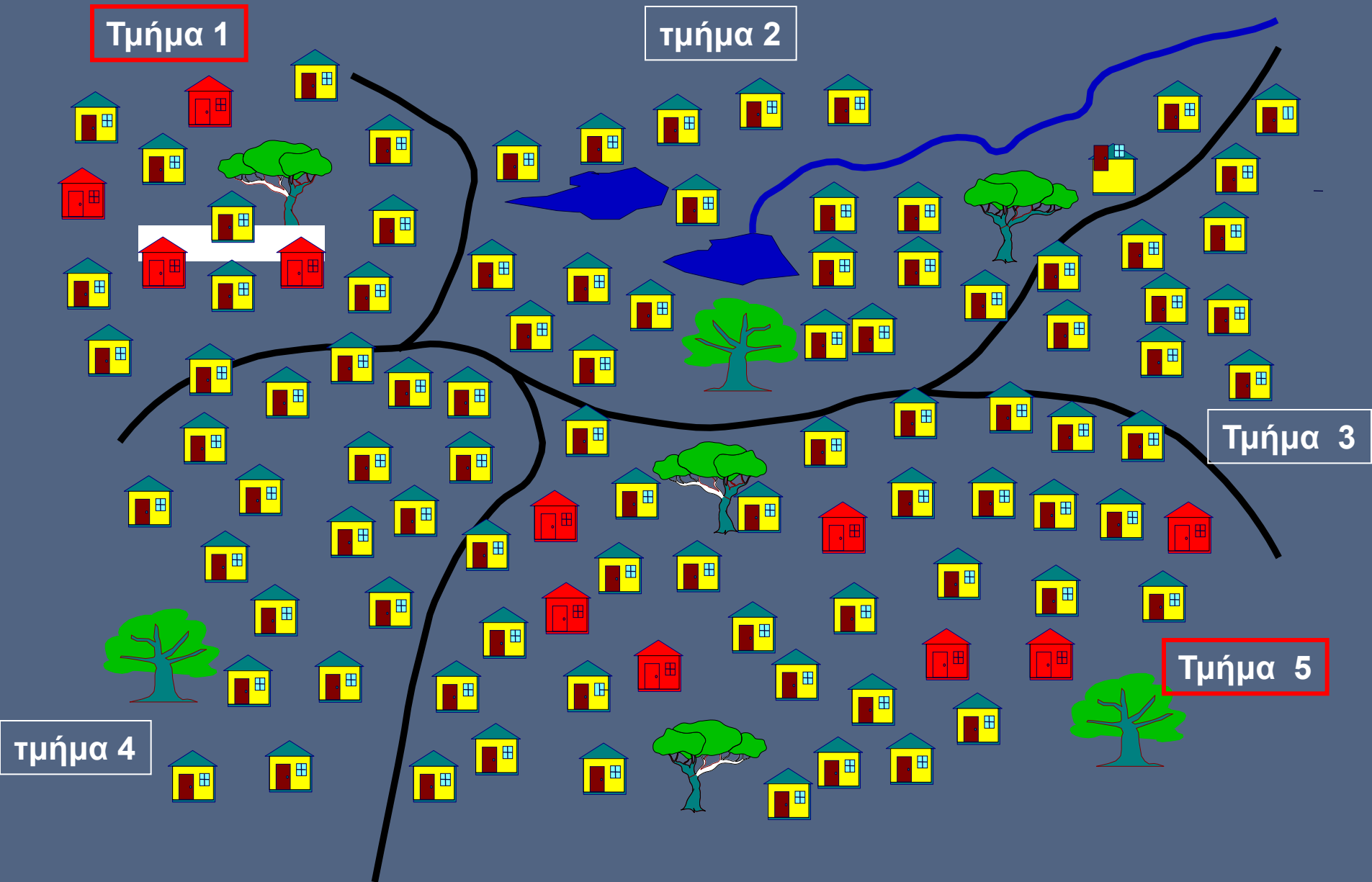
Όλες οι μονάδες δεύτερου επιπέδου περιλαμβάνονται στο δείγμα

Δειγματοληψία ομάδων σε 3 στάδια

Μονάδα δεύτερου Σταδίου => Κατοικίες
Μονάδα τρίτου-σταδίου => Άτομα

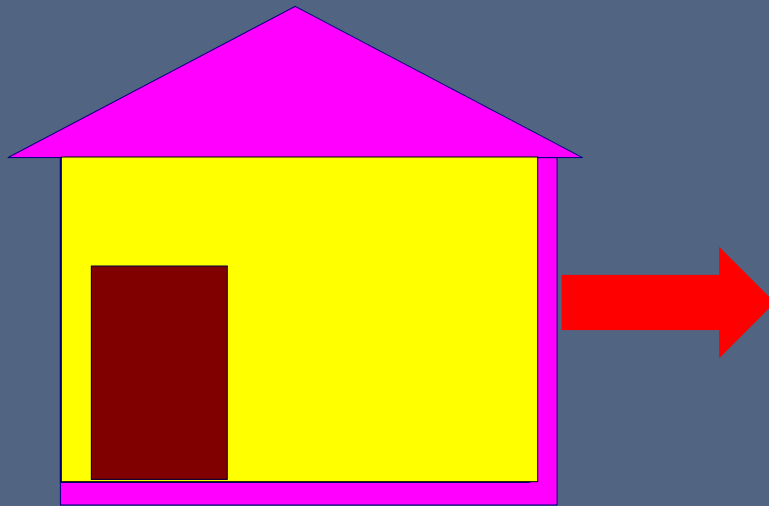


Παράδειγμα: Δειγματοληψία ομάδων σε 3-στάδια



Δειγματοληψία 3-σταδίων

Όλες οι μονάδες του τρίτου σταδίου περιλαμβάνονται στο δείγμα



Δειγματοληψία ομάδων

- **Πλεονεκτήματα**

- Απλή, αν η συμπληρωμένη λίστα των δειγματοληπτικών μονάδων μέσα στον πληθυσμό δεν απαιτείται
- Λιγότερα ταξίδια και χρήματα απαιτούνται

- **Μειονεκτήματα**

- Τα μέλη μιας ομάδας μπορεί να μοιάζουν μεταξύ τους περισσότερο από ότι τα μέλη μιας άλλης ομάδας (ομοιογενή)...και
- Αυτή η “εξάρτηση” χρειάζεται να ληφθεί υπόψη στο μέγεθος δείγματος και στην ανάλυση (“design effect”)

Επιλογή δειγματοληπτικής μεθόδου

- Ο πληθυσμός που θα μελετηθεί
 - Μέγεθος/Γεωγραφική Κατανομή
 - Ετερογένεια σε σεβασμό στην ποικιλότητα
- Διαθέσιμη λίστα δειγματοληπτικών μονάδων
- Επίπεδο ακρίβειας που απαιτείται
- Διαθέσιμα χρήματα

Συμπεράσματα

- Τα δείγματα πιθανότητας είναι τα καλύτερα
- Επιβεβαιώστε
 - Αντιπροσωπευτικότητα
 - Ακρίβεια
-με τους απαραίτητους περιορισμούς

Συμπεράσματα

- Αν αμφιβάλλετε...

Φωνάξτε επιδημιολόγο/στατιστικό !!!!