



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
Τμήμα Βιοϊατρικών Επιστημών

ΕΠΙΔΗΜΙΟΛΟΓΙΑ

Διδάσκουσα: Πρίσκα Ευαγγελία- Ελένη

Διάλεξη 10^η

Ακαδ. έτος 2020-2021 (εαρινό εξάμηνο)

Δειγματοληψία

Σκοπός του μαθήματος

- Ορισμός της δειγματοληψίας
- Γιατί χρησιμοποιούμε δείγμα
- Η έννοια της αντιπροσωπευτικότητας
- Κύριες μέθοδοι δειγματοληψίας
- Σφάλμα δειγματοληψίας
- Υπολογισμός μεγέθους δείγματος

Ορισμός Δειγματοληψίας

«Διαδικασία με την οποία συλλέγονται πληροφορίες από ένα μέρος (υποσύνολο) του συνολικού πληθυσμού αναφοράς»

- Συνήθως, επιλέγουμε δείγμα αντιπροσωπευτικό του πληθυσμού που μας ενδιαφέρει.
- Πληθυσμός: το σύνολο των μονάδων/ατόμων στο οποίο θα αναφέρονται τα ευρήματα της έρευνας (π.χ. γυναικείος πληθυσμός της χώρας ηλικίας 65+ ετών, παιδιά ηλικίας 12 ετών, κτλ.)

Γιατί χρησιμοποιούμε δείγμα;

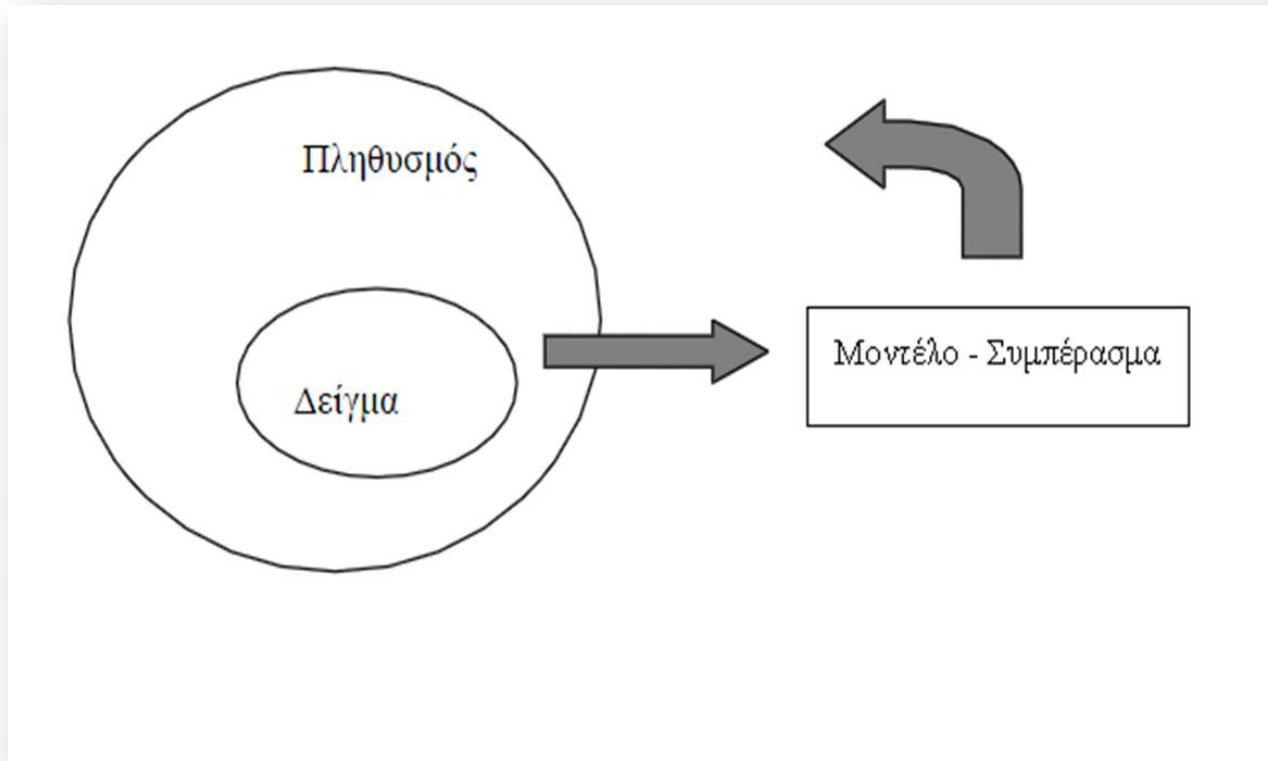
Απογραφή ή δειγματοληψία;;;

Απογραφή: συλλογή στοιχείων από το σύνολο του πληθυσμού.

Δειγματοληψία: συλλογή στοιχείων από υποσύνολο του πληθυσμού με:

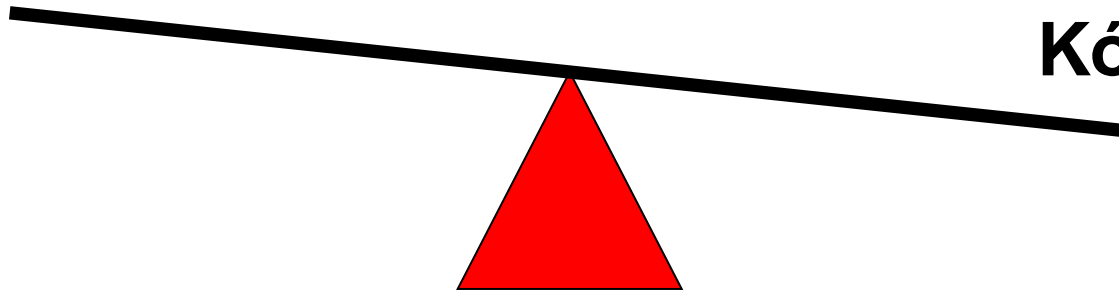
- ελάχιστο κόστος
- μέγιστη ταχύτητα
- αυξημένη ακρίβεια
- βελτιωμένες μέθοδοι εκτίμησης παραμέτρων
- μοναδική μέθοδος έρευνας

Γιατί χρησιμοποιούμε δείγμα;



Δειγματοληψία

Ακρίβεια



Κόστος

- \uparrow ακρίβεια \rightarrow δείγμα \uparrow \rightarrow κόστος \uparrow
- \downarrow ακρίβεια \rightarrow δείγμα \downarrow \rightarrow κόστος \downarrow

Περιγραφή όρων που χρησιμοποιούνται στη δειγματοληψία

Μονάδα δειγματοληψίας

- Τα υπό μελέτη άτομα για τα οποία συλλέγονται πληροφορίες.
 - Παράδειγμα: Παιδιά <5 ετών, εξιτήρια νοσοκομείων, συγκεκριμένα νοσήματα ...

Κλάσμα δειγματοληψίας

- Η αναλογία του μεγέθους του δείγματος προς το σύνολο του πληθυσμού
 - Παράδειγμα: 100 από 2000 (5%)

Περιγραφή όρων που χρησιμοποιούνται στη δειγματοληψία

Πλαίσιο δειγματοληψίας (Sampling frame)

- Κάθε λίστα με όλες τις μονάδες δειγματοληψίας στον πληθυσμό
 - Λίστα των νοικοκυριών, λίστα μονάδων υγείας...

Σχέδιο δειγματοληψίας (Sampling scheme)

- Μέθοδος επιλογής των μονάδων δειγματοληψίας από το πλαίσιο δειγματοληψίας
 - Τυχαία, βολικό δείγμα...

Τι χρειάζεται να ξέρουμε

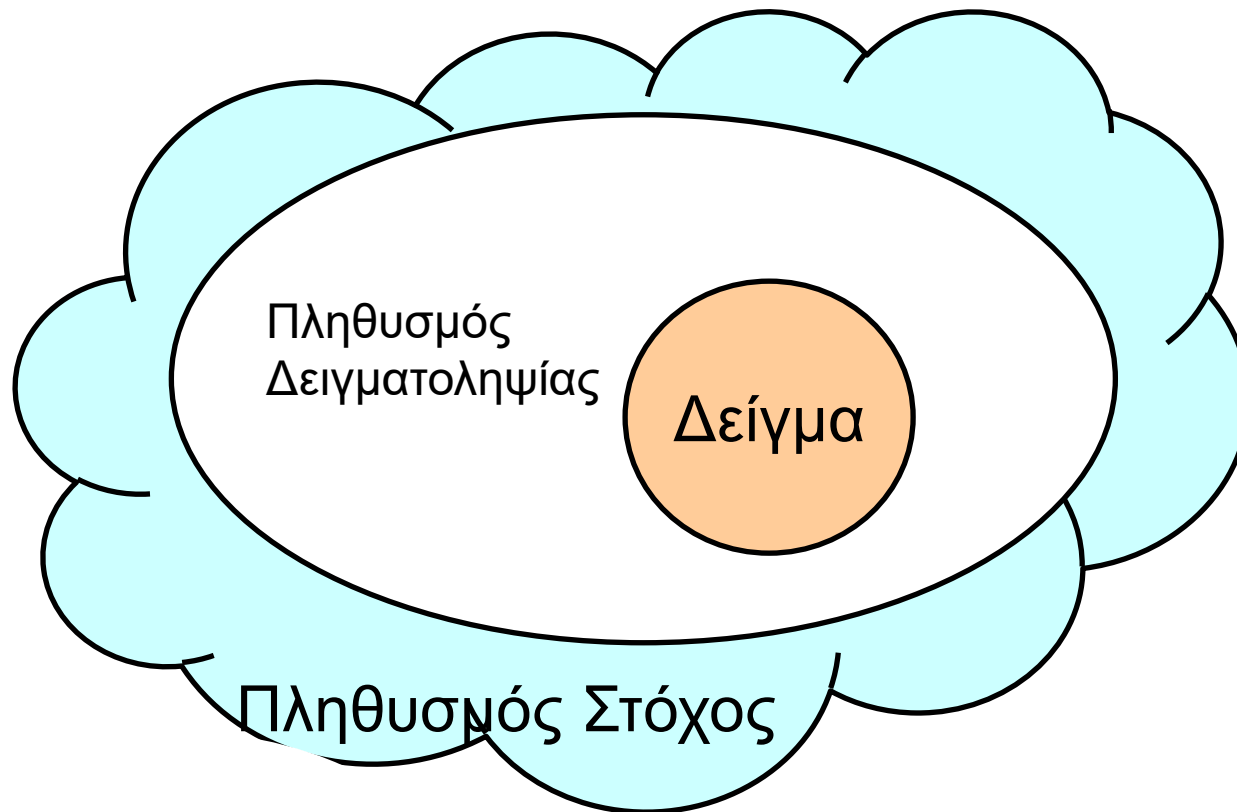
- Έννοιες
 - Αντιπροσωπευτικότητα
 - Μέθοδοι δειγματοληψίας
 - Επιλογή του σωστού σχεδιασμού

- Υπολογισμοί
 - Σφάλμα δειγματοληψίας
 - Επίδραση σχεδιασμού
 - Μέγεθος δείγματος

Σφάλματα

- **Μη δειγματοληπτικά:** μπορούν να συμβούν σε οποιαδήποτε έρευνα
 - Συστηματικά (ή μεροληψίας)
 - Μη συστηματικά
- **Δειγματοληπτικά (δειγματοληψίας):** οφείλονται στη δειγματοληπτική μέθοδο που χρησιμοποιείται (**τυχαία σφάλματα**)
 - Ακρίβειας

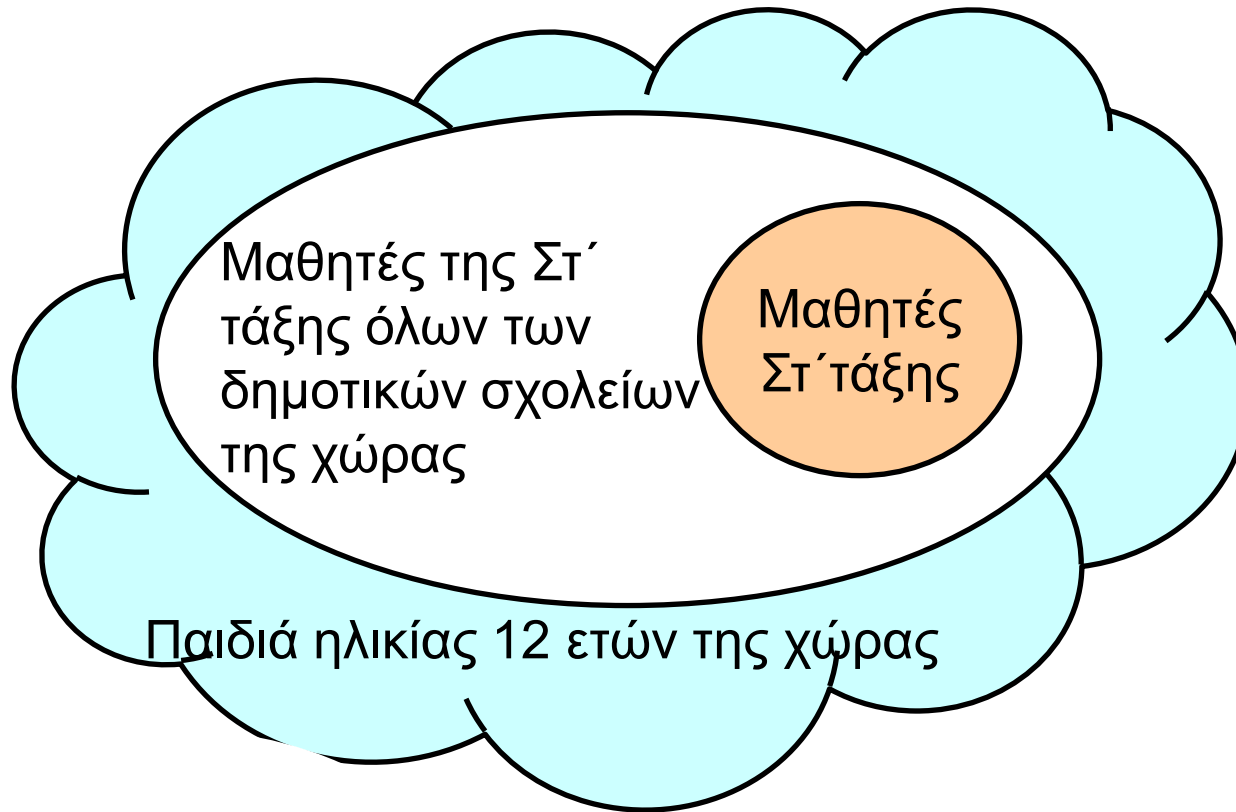
Δειγματοληψία και αντιπροσωπευτικότητα



Πληθυσμός Στόχος → Πληθυσμός Δειγματοληψίας → Δείγμα 12

Δειγματοληψία και αντιπροσωπευτικότητα

Μελέτη επιπολασμού του καπνίσματος σε παιδιά ηλικίας 12 ετών στην Ελλάδα



Πληθυσμός Στόχος → Πληθυσμός Δειγματοληψίας → Δείγμα

Αντιπροσωπευτικότητα

- **Χαρακτηριστικά** - Άτομο
Δημογραφικά χαρακτηριστικά (ηλικία, φύλο,...)
Έκθεση/ευαισθησία
- **Χώρος** - Τόπος (π.χ.: αστικός vs αγροτικός)
- **Χρόνος**
 - Εποχικότητα
 - Ημέρα της εβδομάδας
 - Ώρα στην ημέρα

**Επιβεβαιώστε την αντιπροσωπευτικότητα πριν
αρχίσετε,
επιβεβαιώστε μόλις τελειώσετε!!!!!!**

Μη δειγματοληπτικά σφάλματα

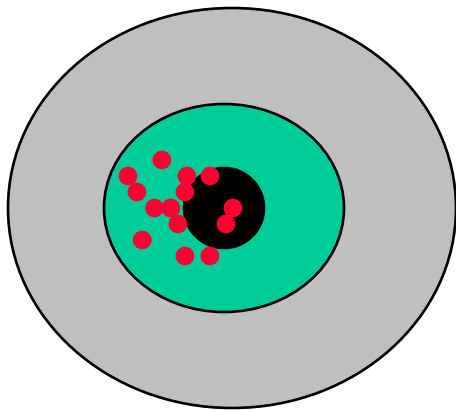
- Συστηματικά
 - Ανακριβείς μετρήσεις
 - κλίμακας (βάρους, ύψους),
π.χ. μια ζυγαριά ζυγίζει πάντα 10 κιλά περισσότερο
 - διαγνωστικά τεστ
- Μη συστηματικά
 - κακώς σχεδιασμένες ερωτήσεις
π.χ. ασαφείς, ελλιπείς
 - σφάλματα συνέντευξης
 - σφάλματα επεξεργασίας δεδομένων

Δειγματοληπτικά σφάλματα

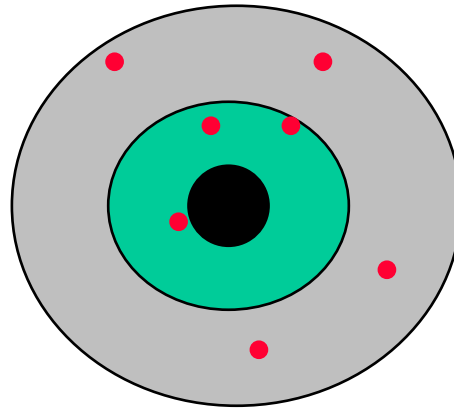
- Κανένα δείγμα δεν είναι ο ακριβής καθρέπτης της εικόνας του πληθυσμού.
- Ακατάλληλη μέθοδος δειγματοληψίας.
- Το μέγεθος του σφάλματος μπορεί να μετρηθεί στη δειγματοληψία με πιθανότητες.
- Εκφράζεται με το τυπικό σφάλμα (standard error)
 - της μέσης τιμής, ποσοστιαία αναλογία, διαφορές, κτλ
- Επίδραση του
 - μεγέθους της διακύμανσης του υπό μέτρηση παράγοντα που μας ενδιαφέρει
 - μεγέθους δείγματος (δείγμα $\uparrow \rightarrow \downarrow$ τυχαίο σφάλμα)

Ποιότητα εκτίμησης παραμέτρων

ακρίβεια
εγκυρότητα

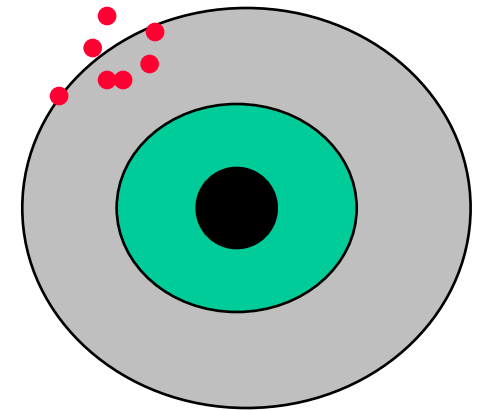


Όχι ακρίβεια



τυχαίο
σφάλμα

ακρίβεια
Όχι εγκυρότητα



συστηματικό
σφάλμα (bias)

Σφάλματα έρευνας: Παράδειγμα

Μέτρηση του ύψους:

- Εάν κρατάνε τη μεζούρα με διαφορετικό τρόπο οι ερευνητές:
 - μειωμένη ακρίβεια
 - τυπικό σφάλμα (standard error)
- Εάν η μεζούρα είναι λάθος (π.χ. +5εκ.)
 - συστηματικό σφάλμα (bias)

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

Τύποι δειγμάτων

- Δείγματα υπολογιζόμενα χωρίς πιθανότητες (Non-probability samples)
- Δείγματα υπολογιζόμενα με πιθανότητες (Probability samples)

Δείγματα υπολογιζόμενα χωρίς πιθανότητες

- Αναλογικά
 - Το δείγμα αντανακλά στη δομή του πληθυσμού
 - Περιορισμοί σε χρόνο και χρήμα
- Βολικά δείγματα
 - Επηηρεασμένο – συστηματικό λάθος (Biased)
 - Το καλύτερο ή χειρότερο σενάριο
 - π.χ. τα πρώτα 100 άτομα που περνούν από ένα σημείο
- Υποκειμενικά δείγματα
 - Επιλογή δείγματος κατά την κρίση κάποιου ειδικού
- Δειγματοληψία χιονοστιβάδας (snowball sampling)
 - Από φίλο σε φίλο, κτλ.

Πιθανότητα το να επιλεγεί μια μονάδα: άγνωστη

Δείγματα με πιθανότητες

- Τυχαία δειγματοληψία.
- Κάθε άτομο έχει μία γνωστή πιθανότητα να επιλεγεί.
- Μειώνει τη δυνατότητα του συστηματικού λάθους επιλογής.
- Επιτρέπει την εφαρμογή της στατιστικής θεωρίας στα αποτελέσματα.

Μέθοδοι που χρησιμοποιούνται στα δείγματα με πιθανότητες

- **Απλή** τυχαία δειγματοληψία (Simple random sampling)
- **Συστηματική** δειγματοληψία (Systematic sampling)
- **Διαστρωματική** δειγματοληψία (Stratified sampling)
- Δειγματοληψία **Συστάδων (καθ' ομάδας)** (Cluster sampling)
- **Πολυσταδιακή** δειγματοληψία (Multistage sampling)

Απλή τυχαία δειγματοληψία

- Βασική αρχή
 - Ίσες πιθανότητες να κληρωθεί η κάθε μονάδα/άτομο.
- Διαδικασία
 - Αρίθμηση όλων των μονάδων
 - Τυχαία κλήρωση αριθμών/μονάδων

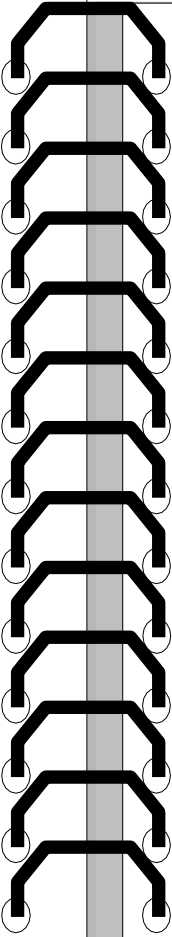
Παράδειγμα Απλής τυχαίας δειγματοληψίας

Να μετρηθεί ο επιπολασμός της τερηδόνας σε 1200 παιδιά που παρακολουθούν το σχολείο

- Λίστα των παιδιών που παρακολουθούν το σχολείο.
- Αριθμημένα παιδιά από το 1 έως το 1200.
- Μέγεθος του δείγματος = 100 παιδιά
- Τυχαία δειγματοληψία 100 νούμερα ανάμεσα στο 1 και το 1200.

Πως θα επιλέξουμε τυχαία;

Απλή τυχαία δειγματοληψία



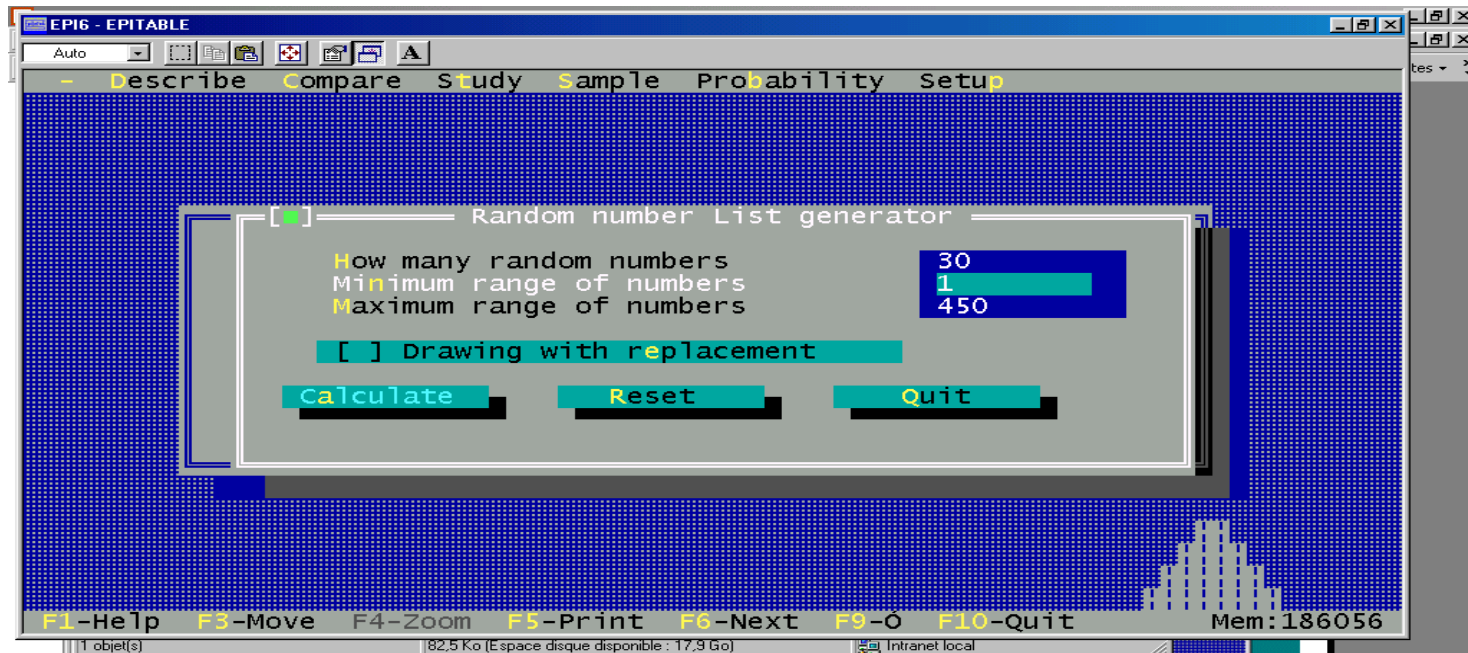
The diagram shows a vertical spinner with 48 numbered slots, each with a pointer. The spinner is divided into two halves by a vertical line. The numbers 1 through 48 are arranged in two columns, with 1-24 on the left and 25-48 on the right. The spinner is shown in a state where the pointer is pointing to the number 20 on the left side.

| | | | |
|----|-----------------|----|----------------|
| 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 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. |

Table of random numbers

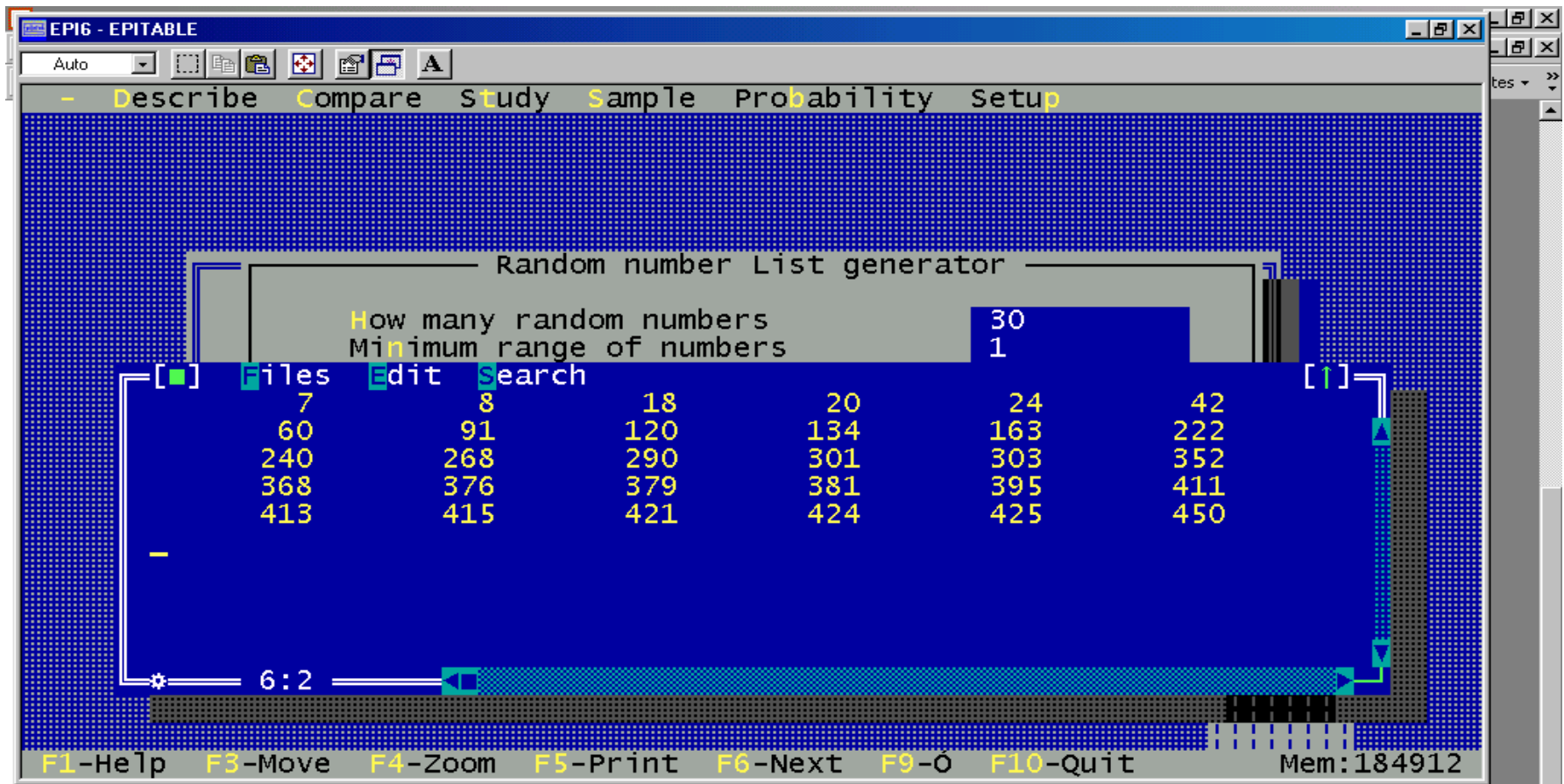
| | | | | | | | |
|------------------|------------------|-------|-------|-------|-------|-------|-------|
| 57172 | 42088 | 70098 | 11333 | 26902 | 29959 | 43909 | 49607 |
| 33883 | 87680 | 28923 | 15659 | 09839 | 45817 | 89405 | 70743 |
| 77950 | 67344 | 10609 | 87119 | 15859 | 74577 | 42791 | 75889 |
| 11607 | 11596 | 01796 | 24498 | 17009 | 67119 | 00614 | 49529 |
| 56149 | 55678 | 38169 | 47228 | 49931 | 94303 | 67448 | 31286 |
| 80719 | 65101 | 77729 | 83949 | 83358 | 75230 | 56624 | 27549 |
| 93809 | 19505 | 82000 | 79068 | 45552 | 86776 | 48980 | 56684 |
| 40950 | 86216 | 48161 | 17646 | 24164 | 35513 | 94057 | 51834 |
| 12182 | 59744 | 65695 | 83710 | 41125 | 14291 | 74773 | 66391 |
| 13382 | 48076 | 73151 | 48724 | 35670 | 38453 | 63154 | 58116 |
| 38629 | 94576 | 48859 | 75654 | 17152 | 66516 | 78796 | 73099 |
| 60728 | 32063 | 12431 | 23898 | 23683 | 10853 | 04038 | 75246 |
| 01881 | 99056 | 46747 | 08846 | 01331 | 88163 | 74462 | 14551 |
| 23094 | 29831 | 95387 | 23917 | 07421 | 97869 | 88092 | 72201 |
| 15243 | 21100 | 48125 | 05243 | 16181 | 39641 | 36970 | 99522 |
| 53501 | 58431 | 68149 | 25405 | 23463 | 49168 | 02048 | 31522 |
| 07698 | 24181 | 01161 | 01527 | 17046 | 31460 | 91507 | 16050 |
| 22921 | 25930 | 79579 | 43488 | 13211 | 71120 | 91715 | 49881 |
| 68127 | 00501 | 37484 | 99278 | 28751 | 80855 | 02035 | 10910 |
| 55309 | 10713 | 36439 | 65660 | 72554 | 77021 | 46279 | 22705 |
| 92034 | 90892 | 69853 | 06175 | 61221 | 76825 | 18239 | 47687 |
| 50612 | 84077 | 41387 | 54107 | 09190 | 74305 | 68196 | 75634 |
| 81415 | 98504 | 32168 | 17822 | 49946 | 37545 | 47201 | 85224 |
| 38461 | 44528 | 30953 | 08633 | 08049 | 68698 | 08759 | 45611 |
| 07556 | 24587 | 88753 | 71626 | 64864 | 54986 | 38964 | 83534 |
| 60557 | 50031 | 75829 | 05622 | 30237 | 77795 | 41870 | 26300 |

EPITABLE: δημιουργία λίστας με τυχαίους αριθμούς



Μπορεί να δημιουργηθεί και στο Excel: συνάρτηση RANDBETWEEN (bottom;top).

EPITABLE: κατάλογος τυχαίων αριθμών



Μπορεί να δημιουργηθεί και στο Excel.

Απλή τυχαία δειγματοληψία

- Πλεονεκτήματα
 - Απλή εφαρμογή.
 - Σφάλμα δειγματοληψίας μπορεί να μετρηθεί εύκολα.
- Μειονεκτήματα
 - Ανάγκη ολοκληρωμένης λίστας μονάδων.
 - Δεν καταφέρνει πάντοτε την καλύτερη αντιπροσωπευτικότητα.
 - Οι μονάδες μπορεί να είναι διάσπαρτες.

Συστηματική δειγματοληψία

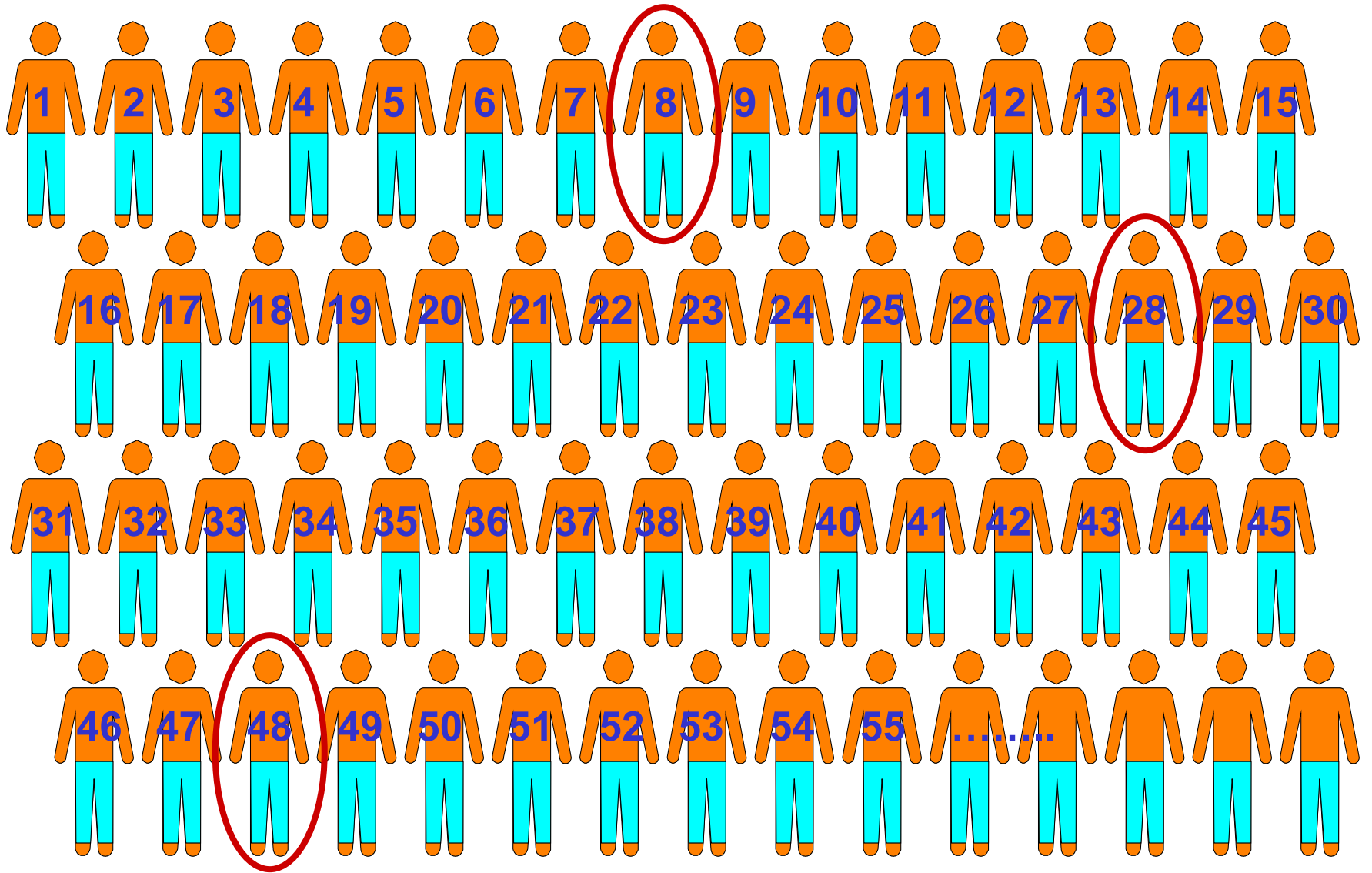
- Βασική αρχή
 - Επιλογή μονάδων ανά **διάστημα δειγματοληψίας** (σε τακτά διαστήματα).
- Διαδικασία
 - Ταξινόμηση μονάδων σε λίστα.
 - Υπολογισμός «**διάστημα δειγματοληψίας**»: $k=N/n$
 - Επιλογή τυχαίου σημείου εκκίνησης (τυχαίος αριθμός μεταξύ 1 και k).
 - Επιλογή μονάδων ανά διάστημα δειγματοληψίας.

Παράδειγμα Συστηματικής δειγματοληψίας

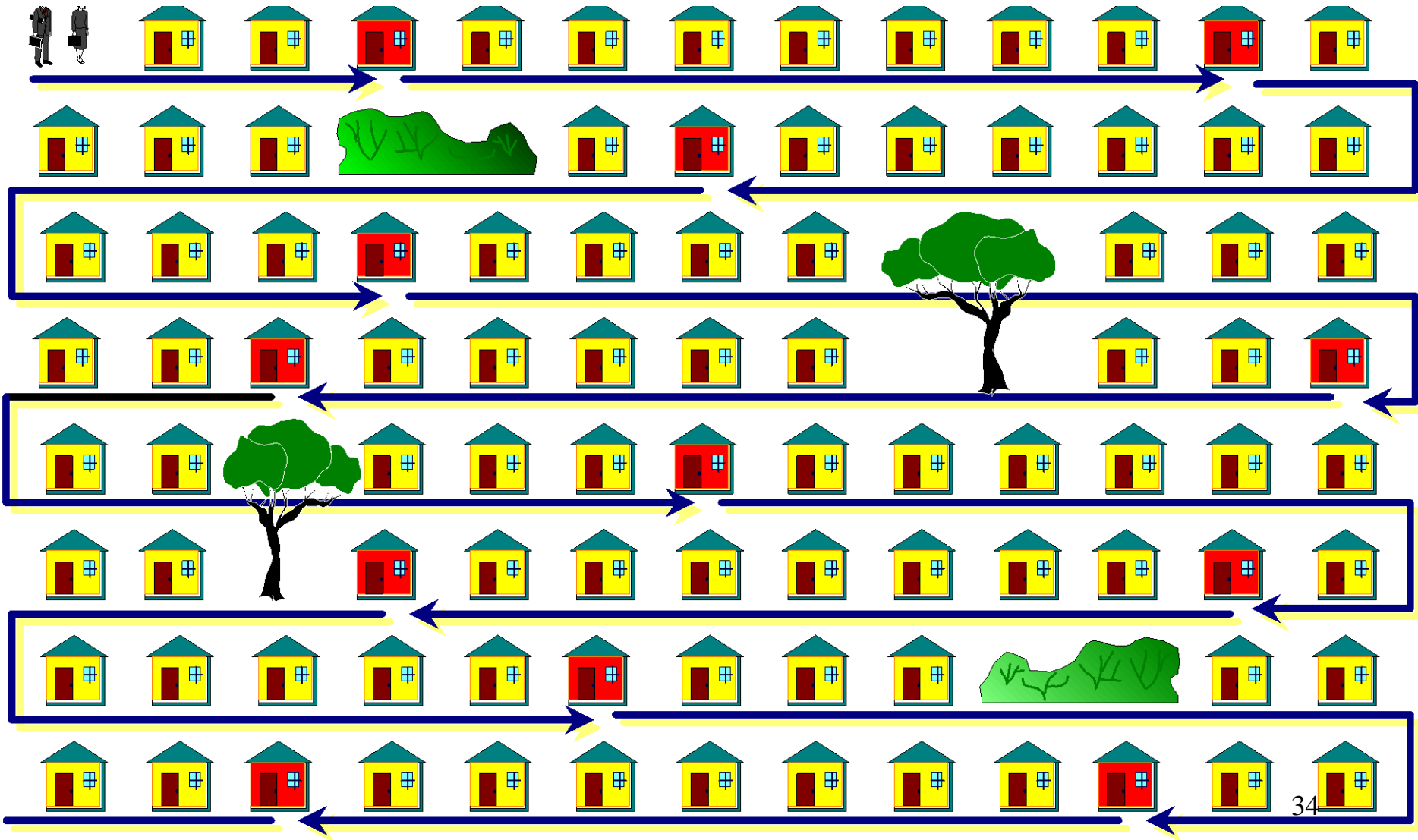
- $N = 1200$, και δείγμα $n = 60$,
⇒ διάστημα δειγματοληψίας:
→ $1200/60 = 20$
- Κατάλογος ατόμων από το 1 μέχρι το 1200
- Επιλέξτε τυχαία ένα αριθμό ανάμεσα στο 1 και στο 20 (πχ.: 8)
⇒ 1ο άτομο που επιλέγεται = το 8ο στον κατάλογο
⇒ 2ο άτομο = $8 + 20 =$ το 28ο
(48ο, 68ο, κτλ.....)

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. |



Συστηματική δειγματοληψία



Συστηματική δειγματοληψία

- Πλεονεκτήματα
 - Απλή εφαρμογή.
 - Πιο αντιπροσωπευτικό της απλής τυχαίας δειγματοληψίας (αν δεν υπάρχει πρόβλημα περιοδικότητας).
- Μειονεκτήματα
 - Ανάγκη ολοκληρωμένης λίστας μονάδων.
 - Δεν καταφέρνει πάντοτε την καλύτερη αντιπροσωπευτικότητα.
 - Πρόβλημα περιοδικότητας στην λίστα.

Διαστρωματική (Στρωματοποιημένη) δειγματοληψία

Χρησιμοποιείται όταν ο πληθυσμός μπορεί να διαιρεθεί σε στρώματα (ομάδες) βάσει ενός χαρακτηριστικού και αναμένονται «μεγάλες» διαφορές μεταξύ τους.

- Βασική αρχή
 - Επιλογή μονάδων ξεχωριστά σε κάθε στρώμα/ομάδα.
- Διαδικασία:
 - Ταξινόμηση του πληθυσμού σε **ομοιογενή** στρώματα
 - π.χ. νομοί, αστικότητα, υψόμετρο περιοχής, ηλικιακή ομάδα, τμήμα φοίτησης
 - Επιλογή μονάδων από κάθε στρώμα (απλή ή συστηματική δειγματοληψία)
 - Μέγεθος δείγματος
 - Ανάλογο του πληθυσμού κάθε στρώματος
 - Ίδιο από κάθε στρώμα
 - Συνδέστε τα αποτελέσματα όλων των στρωμάτων.

Παράδειγμα Διαστρωματικής δειγματοληψίας

- Υπολογισμός της εμβολιαστικής κάλυψης σε μια χώρα, έστω Ελλάδα...
- Επιλέγεται ένα δείγμα σε κάθε στρώμα/περιοχή (stratum).
- Γίνονται υπολογισμοί σε κάθε στρώμα/περιοχή.
- Δίνεται συντελεστής βαρύτητας σε κάθε στρώμα, ώστε να υπολογίσουμε την εμβολιαστική κάλυψη στο σύνολο της χώρας.

Παράδειγμα Διαστρωματικής δειγματοληψίας

Επιλογή δείγματος ($n=1000$) ανάλογο του μεγέθους του πληθυσμού

| Περιοχή | Πληθυσμός | Αναλογία | Δείγμα | Κλάσμα Δειγματοληψίας |
|---------------|--------------|----------|-------------------------|--------------------------|
| Αστική | 7000 | 70% | $1000 \times 0,7 = 700$ | 10 % |
| Αγροτική | 3000 | 30% | $1000 \times 0,3 = 300$ | 10 % |
| Σύνολο | 10000 | | | 1000 |

(n/N)

Θα μπορούσε να επιλεγθούν και 500 άτομα από
κάθε στρώμα!

Διαστρωματική δειγματοληψία

- Πλεονεκτήματα

- Πιο ακριβής εάν η μεταβλητή συνδέεται με τα στρώματα (strata).
- Καλύτερη αντιπροσώπευση του πληθυσμού.
- Όλες οι υποομάδες αντιπροσωπεύονται, επιτρέποντας ξεχωριστά συμπεράσματα για κάθε μια από αυτές.

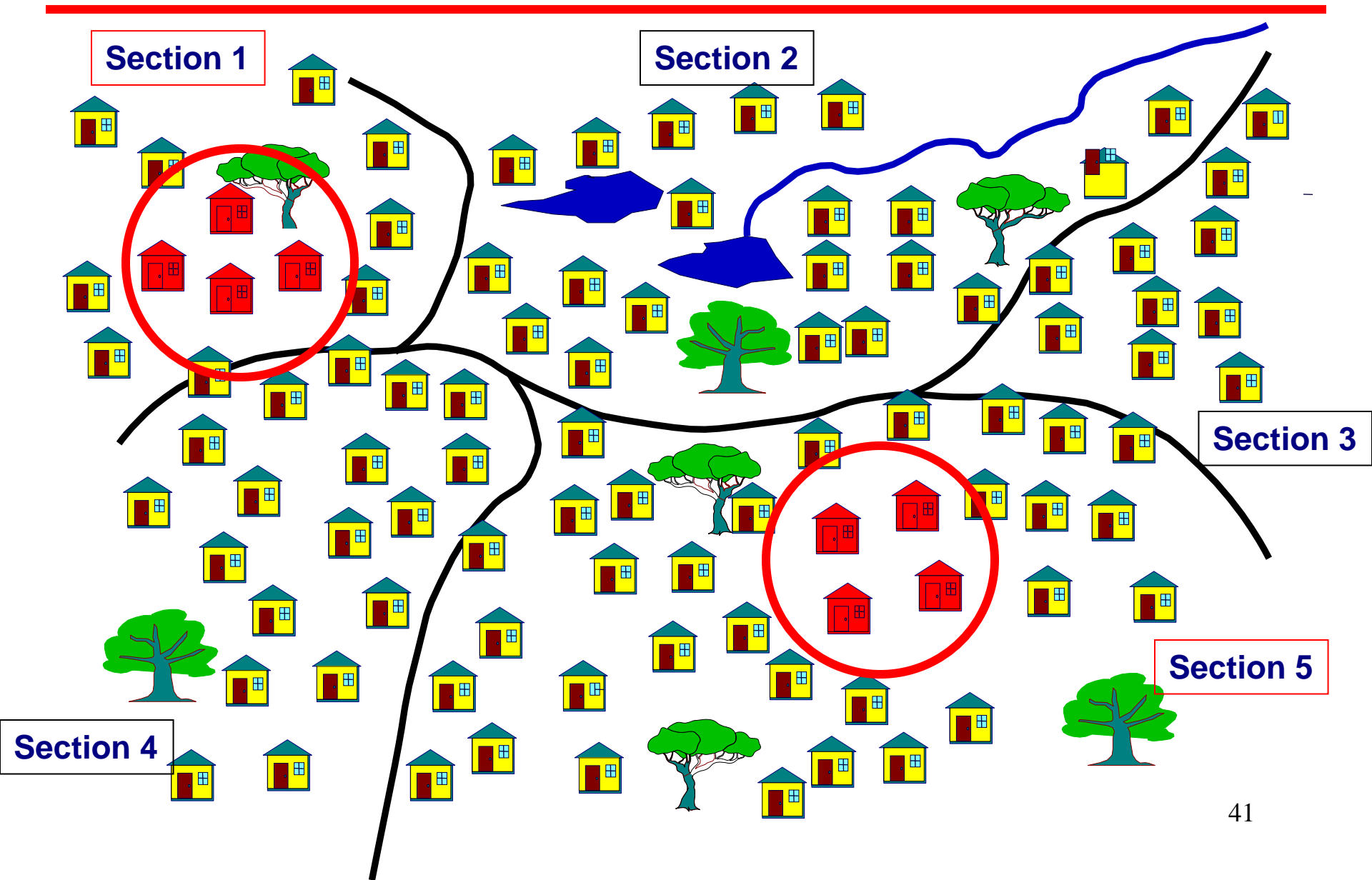
- Μειονεκτήματα

- Το σφάλμα δειγματοληψίας είναι δύσκολο να μετρηθεί.
- Απώλεια της ακρίβειας εάν πολύ μικρά νούμερα περιληφθούν στο δείγμα από τα επιμέρους στρώματα (strata).

Δειγματοληψία Συστάδων (Cluster sampling)

- Βασική αρχή:
 - Επιλογή συστάδων (αντί μονάδων)
 - Συστάδες (clusters) = περιλαμβάνουν πολλές μονάδες
- Διαδικασία
 - Ταξινόμηση του πληθυσμού σε συστάδες.
 - Επιλογή συστάδων (π.χ. σχολεία, συνοικίες, πολυκατοικίες, οικοδομικά τετράγωνα) με:
 - Απλή τυχαία δειγματοληψία
 - (Συστηματική) δειγματοληψία με Πιθανότητα επιλογής Ανάλογη του Μεγέθους (probability proportional to size: PPS)
 - Στις επιλεγθείσες συστάδες, όλες οι μονάδες ή τυχαίο δείγμα αυτών περιλαμβάνονται στο δείγμα.

Παράδειγμα Δειγματοληψίας Συστάδων (Cluster sampling)



Δειγματοληψία Συστάδων

- Πλεονεκτήματα
 - Είναι απλή γιατί δεν χρειάζεται η δημιουργία καταλόγου των μονάδων δειγματοληψίας του πληθυσμού.
 - Λιγότερο κόστος και χρόνος (ταξίδια) για την ολοκλήρωσή της.
- Μειονεκτήματα
 - Ανακριβής, αν οι ομάδες είναι ανομοιογενείς και έτσι η διακύμανση του δείγματος είναι μεγαλύτερη από τη διακύμανση του πληθυσμού (μεγάλη επίδραση σχεδιασμού).
 - Δυσκολία στη μέτρηση του σφάλματος δειγματοληψίας.

Παράδειγμα δειγματοληψίας Συστάδων

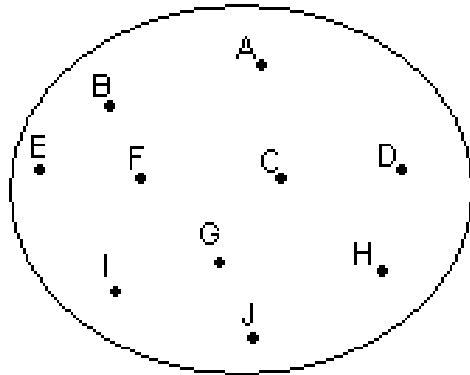
Αξιολόγηση της εμβολιαστικής κάλυψης:

- Χωρίς κατάλογο ατόμων
- Συνολικός πληθυσμός
- Επιλέξτε τυχαία 30 ομάδες
- 30 ομάδες των 7 παιδιών = 210 παιδιά

Επιλογή συστάδων/ομάδων (clusters)

Χρειαζόμαστε :

- Χάρτη της περιοχής.
- Κατανομή του πληθυσμού ανά χωριό ή περιοχή.
- Κατανομή του πληθυσμού ανά ηλικία (12-23 μηνών: 3%).

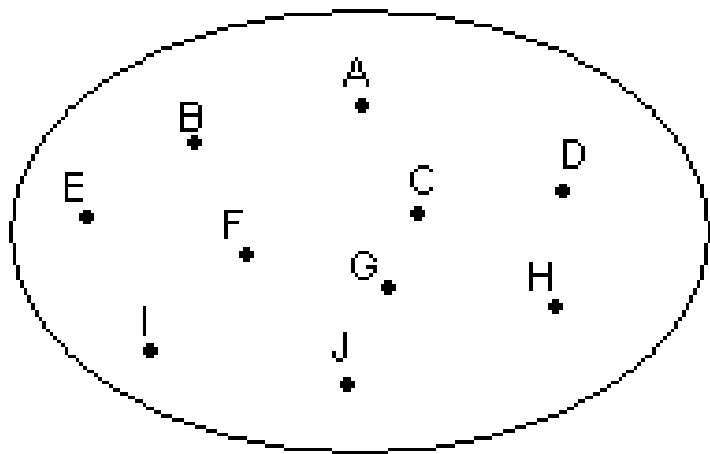


| Χωριό | Πληθ. | 12-23 μήνες |
|-------|--------|-------------|
| A | 53000 | 1600 |
| B | 7300 | 220 |
| C | 106000 | 3200 |
| D | 13000 | 400 |
| E | 26500 | 800 |
| F | 6600 | 200 |
| G | 40000 | 1200 |
| H | 6600 | 200 |
| I | 53000 | 1600 |
| J | 13200 | 400 |

Κατανομή των συστάδων

Συστηματική δειγματοληψία με πιθανότητα ανάλογη του μεγέθους

1. Υπολογίστε αθροιστικά τον πληθυσμό



| | | |
|---|------|------|
| A | 1600 | 1600 |
| B | 220 | 1820 |
| C | 3200 | 5020 |
| D | 400 | 5420 |
| E | 800 | 6220 |
| F | 200 | 6420 |
| G | 1200 | 7620 |
| H | 200 | 7820 |
| I | 1600 | 9420 |
| J | 400 | 9820 |

Σύνολο πληθυσμού = 9820

Κατανομή των συστάδων

Συστηματική δειγματοληψία με πιθανότητα ανάλογη του μεγέθους

2. Υπολογίστε το διάστημα δειγματοληψίας:

$$K = \frac{9820}{30} = 327$$

Κληρώστε ένα τυχαίο νούμερο (ανάμεσα στο 1 και στο 327)

Παράδειγμα: 62

Ξεκινήστε από το χωριό Α στο οποίο περιλαμβάνεται το “62” και κληρώστε τις ομάδες προσθέτοντας το διάστημα δειγματοληψίας

| | | |
|---|------|--|
| A | 1600 | |
| B | 1820 | |
| C | 5020 | |
| D | 5420 | |
| E | 6220 | |
| F | 6420 | |
| G | 7620 | |
| H | 7820 | |
| I | 9420 | |
| J | 9820 | |

Κατανομή των συστάδων

Συστηματική δειγματοληψία με πιθανότητα ανάλογη του μεγέθους

3. Επιλογή - Κλήρωση νοικοκυριών και παιδιών

Στο σημείο

- Πηγαίνετε στο κέντρο του χωριού, επιλέξτε τυχαία κατεύθυνση
- Αριθμήστε τα σπίτια σε αυτή την κατεύθυνση
 - π.χ.: 21
- Κληρώστε ένα τυχαίο νούμερο (ανάμεσα στο 1 και στο 21) για να βρείτε το σπίτι που θα επισκεφθείτε
- Σε αυτό το σπίτι προχωρήστε μέχρι να βρείτε 7 παιδιά

Πολυσταδιακή Δειγματοληψία

- Βασική αρχή:
 - Επιλογή δείγματος σε πολλαπλά στάδια (διαδοχικές δειγματοληψίες)
 - Σύνθετη μορφή δειγματοληψίας συστάδων
- **Παράδειγμα :**
Μονάδα δειγματοληψίας = νοικοκυριό
 - 1ο στάδιο: επιλογή περιοχών/συνοικιών
 - 2ο στάδιο: επιλογή οικοδομικών τετραγώνων
 - 3ο στάδιο: επιλογή πολυκατοικιών
 - 4ο στάδιο: επιλογή νοικοκυριών

Παράδειγμα Πολυσταδιακής δειγματοληψίας

Υπολογισμός εμβολιαστικής κάλυψης παιδιών 7
ετών στην Ελλάδα

Μονάδα δειγματοληψίας: τμήμα Α΄ τάξης Δημοτικού

- 1ο στάδιο: επιλογή δημοτικών σχολείων
- 2ο στάδιο: επιλογή τμημάτων Α΄ τάξης
- (3ο στάδιο: επιλογή μαθητών Α΄ τάξης)

Επίδραση σχεδιασμού Δειγματοληψία Συστάδων

Global variance

$$\text{Var srs} = \frac{p(1-p)}{n}$$

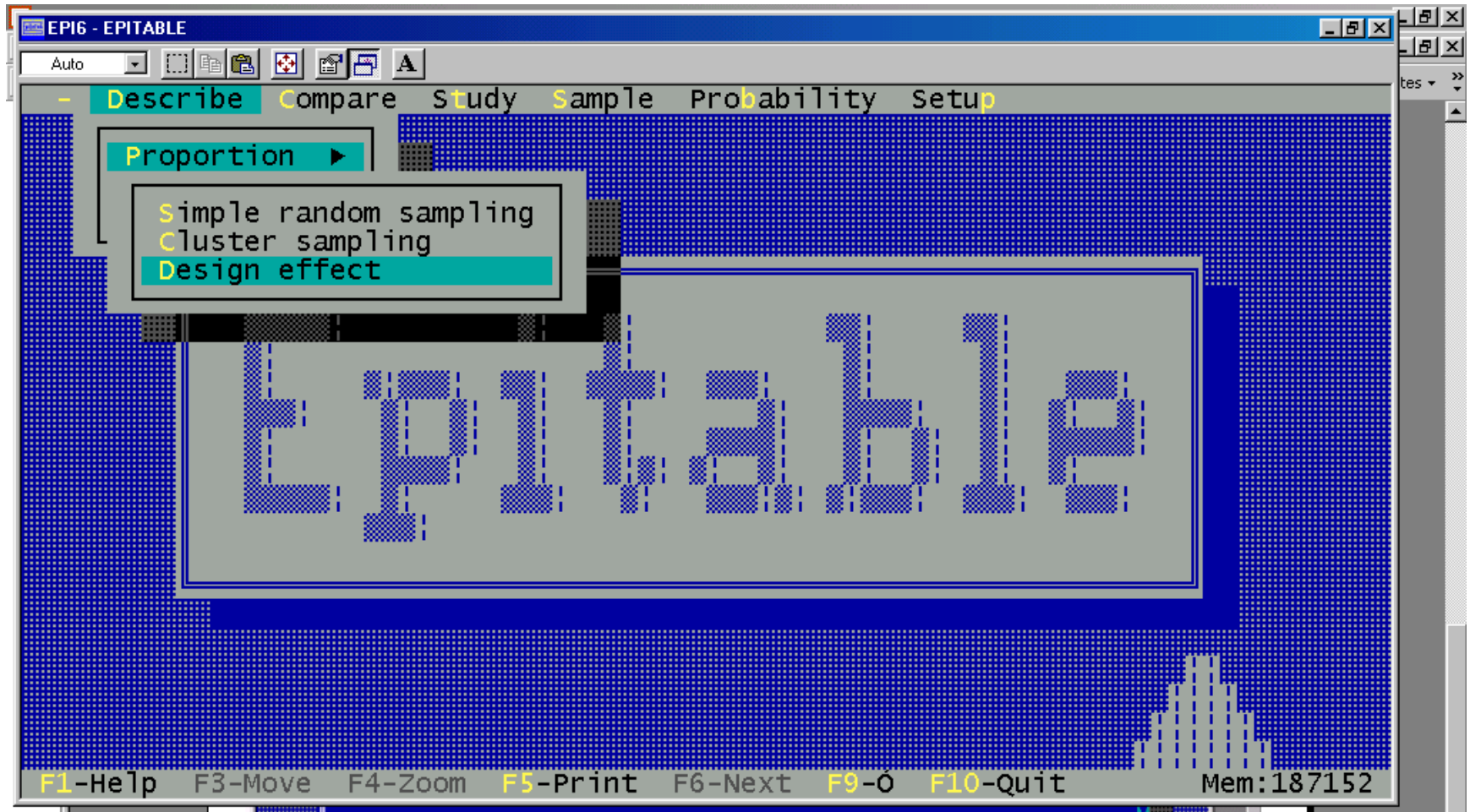
Cluster variance

$$\text{Var clus} = \frac{\sum (p_i - p)^2}{k(k-1)}$$

$$\text{Design effect} = \frac{\text{Var clus}}{\text{Var srs}}$$

p= global proportion
p_i= proportion in each stratum
n= number of subjects
k= number of strata
srs= simple random sampling

EPITABLE: Υπολογισμός επίδρασης σχεδιασμού



Ειδική μέθοδος Δειγματοληψίας

Lot Quality Assurance Sampling

Προέλευση LQAS

- Bell Laboratories (Ηλεκτρικές συσκευές):
 - Έλεγχος του χρόνου λειτουργίας των ηλεκτρικών λαμπτήρων
 - Εξ ορισμού κατεστραμμένες;;;
- Ποιοτικός έλεγχος
 - Ελάχιστο δείγμα
 - Μεγάλη ασφάλεια

Πως χρησιμοποιούμε το LQAS ?

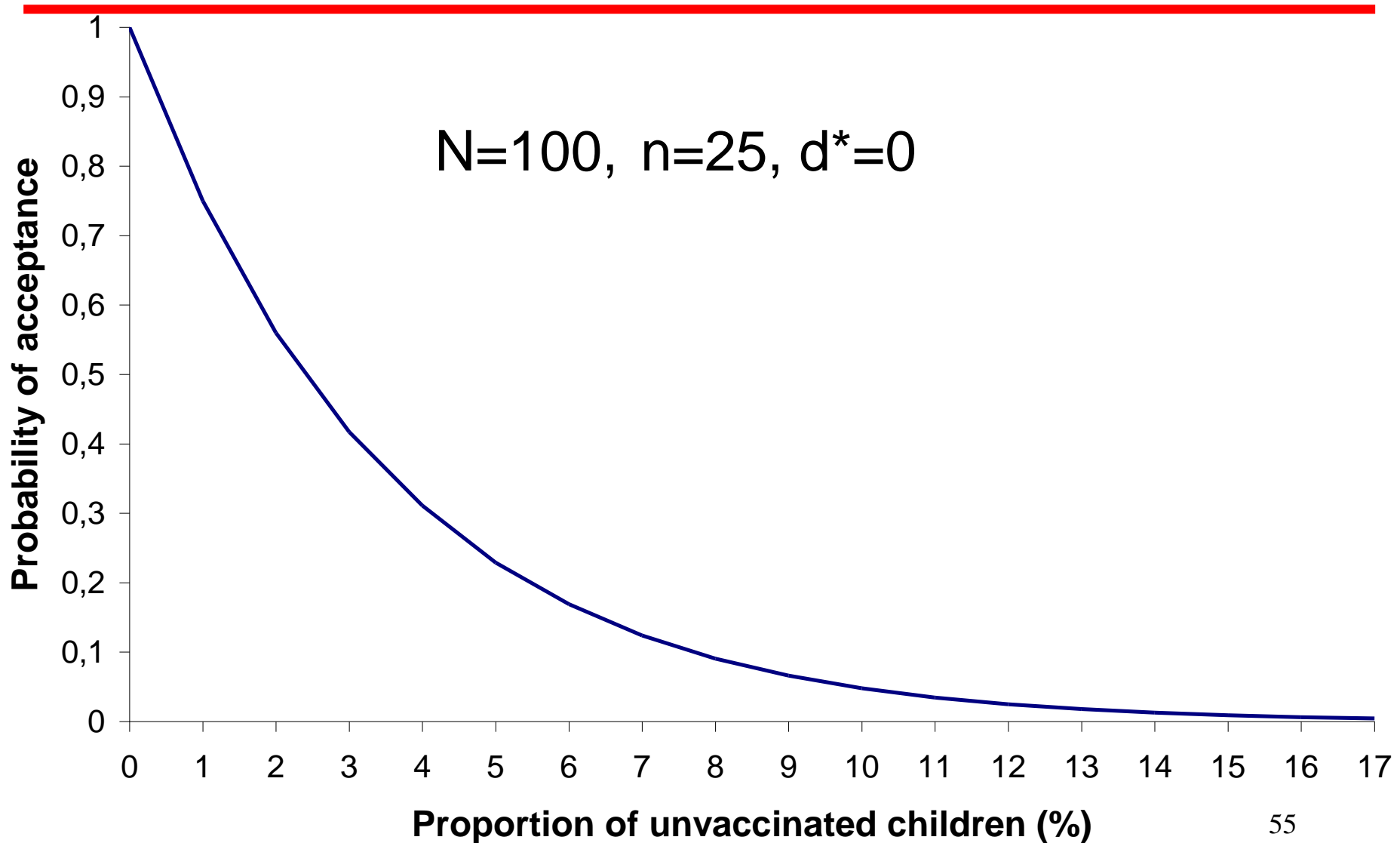
- Ορίζουμε τη σειρά παραγωγής (production)
 - Παρτίδα/σωρός (lot)
- Κατάταξη των σειρών παραγωγής σε αποδεκτές και μη αποδεκτές.
- Σχετικά μικρό τυχαίο δείγμα.
- Μετρούμε τις αποτυχίες “defectives” (binary variable)
 - If # defectives > specified #, reject
 - If # defectives \leq specified #, accept

Παράδειγμα: 100 λαμπτήρες

Υπόθεση: $<10\%$ των λαμπτήρων είναι ελαττωματικοί

- 25 λαμπτήρες επιλέχθηκαν τυχαία
- Υπολογισμός των ελαττωματικών
 - Εάν >0 ελαττωματικές \Rightarrow απόρριψη
 - Εάν καμία ελαττωματική \Rightarrow αποδοχή

Operational characteristic curve

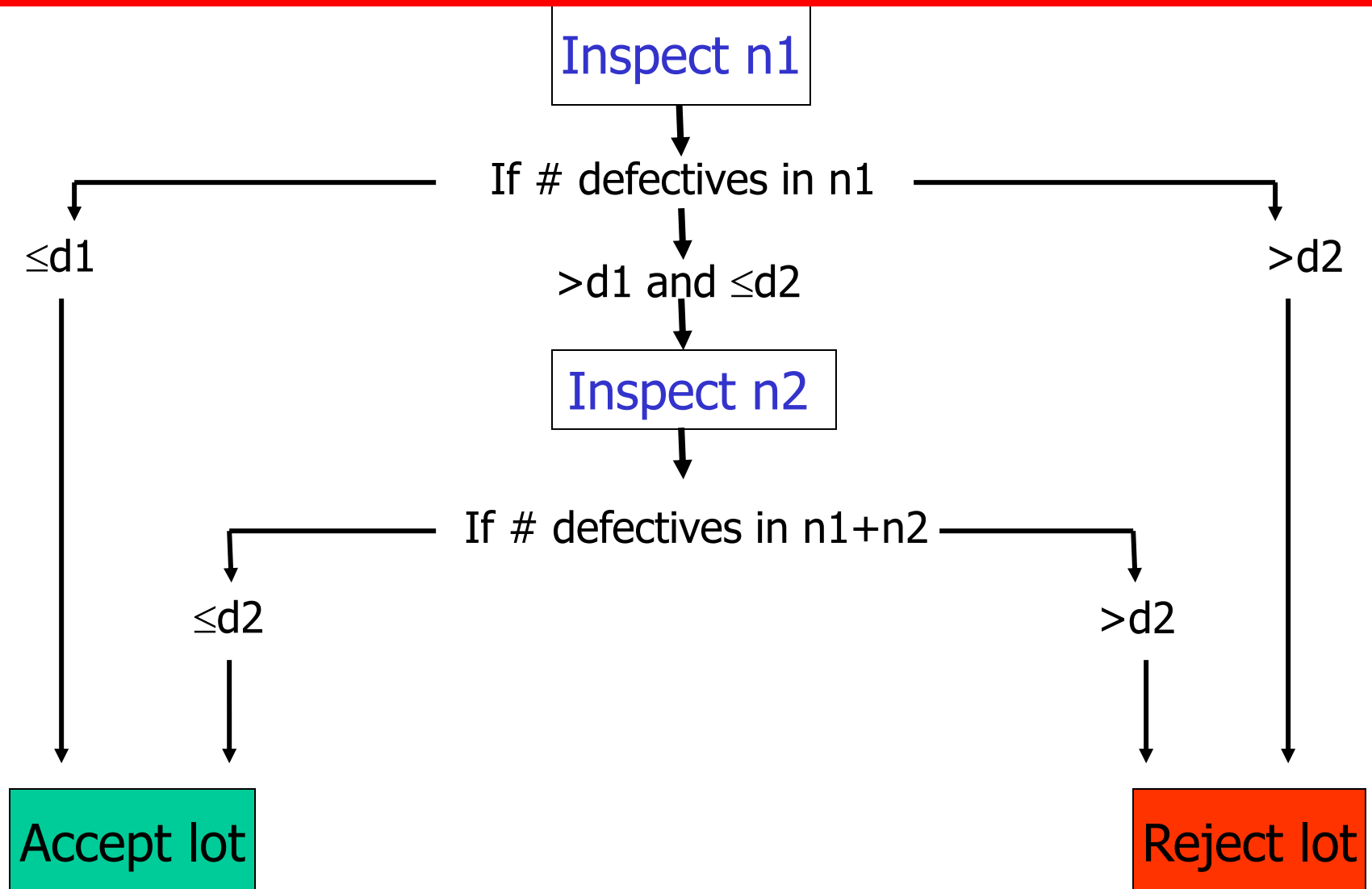


Μονή ή διπλή δειγματοληψία?

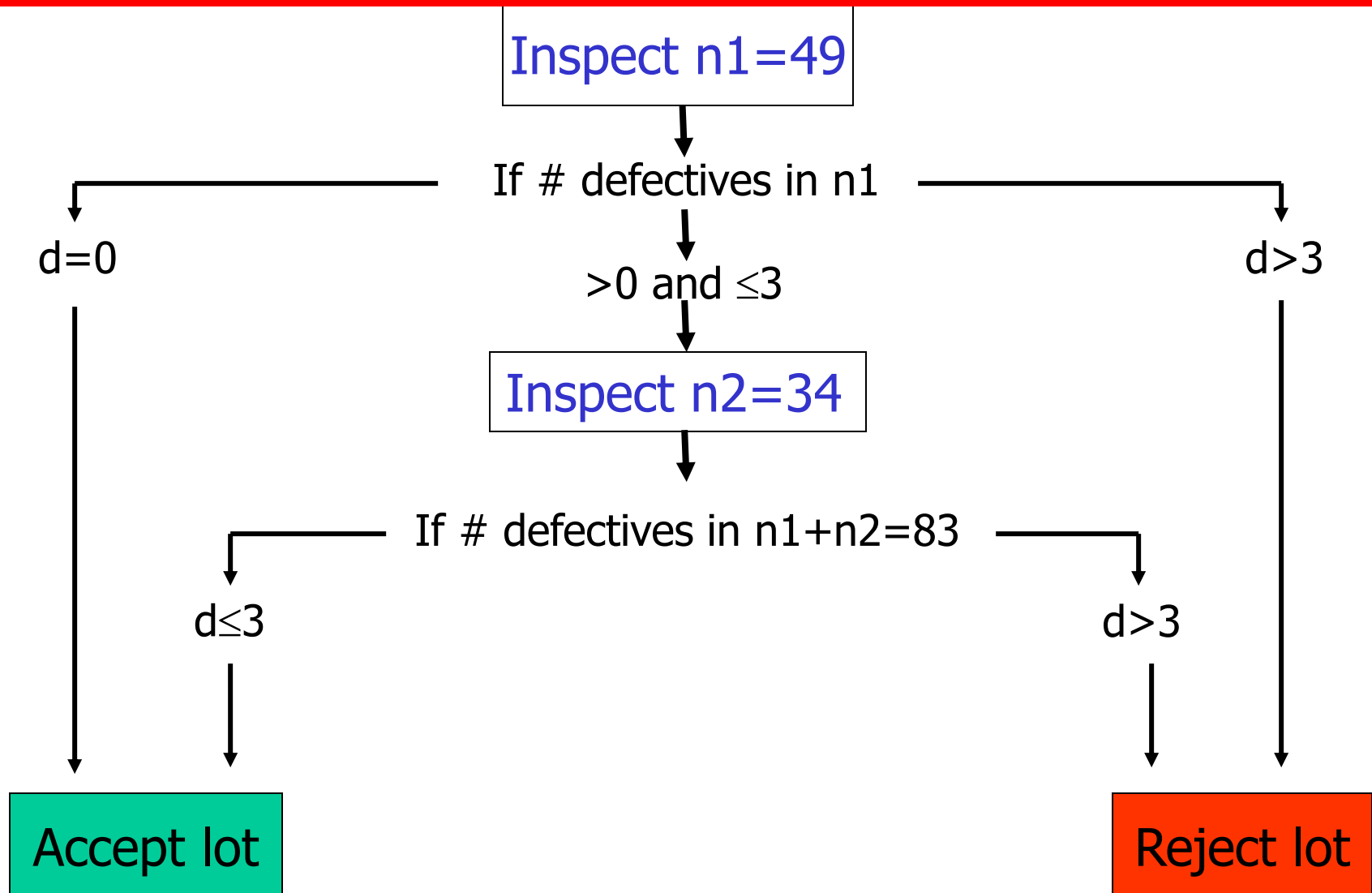
- Μονή δειγματοληψία
 - Απέτυχε να περάσει.
- Διπλή δειγματοληψία
 - Συνήθως, πρώτο δείγμα μικρότερο.
 - Δεύτερο δείγμα, εφόσον με το πρώτο δεν μπορούμε να αποφασίσουμε.
 - Για σπάνια γεγονότα.

Παράδειγμα: $n_1=49$ $d_1=0$; $n_1+n_2=83$ $d_2=3$

Διπλή δειγματοληψία



Διπλή δειγματοληψία



Υπολογισμός του μεγέθους του δείγματος σε LQAS

- Τι χαρακτηριστικό μετρούμε?
- Πως ορίζεται η σειρά παραγωγής?
- Ποια είναι η αναμενόμενη συχνότητα του παράγοντα(P_a)?
- Ποιο είναι το αποδεκτό όριο P_o ?
- Μονή ή διπλή δειγματοληψία?
- Πόσο μεγάλα είναι το α και β ?
- Υπολογισμός του μεγέθους του δείγματος και του d^* με πίνακες H/Y.

LQAS (8)

Table 9-2 LQAS sample sizes at the household and retail levels for various values of P_a and P_o .†

| P_a | P_o | n | d^* | P_a | P_o | n | d^* | P_a | P_o | n | d^* |
|-------|-------|-----|-------|-------|-------|----|-------|-------|-------|----|-------|
| 99% | 95% | 123 | 2 | 90% | 80% | 83 | 10 | 80% | 65% | 56 | 13 |
| | 90% | 42 | 1 | | 75% | 42 | 5 | | 60% | 33 | 8 |
| | 85% | 23 | 0 | | 70% | 26 | 3 | | 55% | 22 | 5 |
| | 80% | 16 | 0 | | 65% | 18 | 2 | | 50% | 15 | 4 |
| | 75% | 11 | 0 | | 60% | 13 | 2 | | 45% | 11 | 3 |
| | 70% | 9 | 0 | | 55% | 10 | 1 | | 40% | 9 | 2 |
| | 65% | 7 | 0 | | 50% | 8 | 1 | | 35% | 7 | 2 |
| | 60% | 6 | 0 | | 45% | 6 | 1 | | 30% | 5 | 1 |
| | | | | 40% | 5 | 1 | 25% | 4 | 1 | | |
| 95% | 85% | 60 | 4 | 85% | 70% | 50 | 9 | 75% | 55% | 61 | 18 |
| | 80% | 32 | 2 | | 65% | 30 | 6 | | 50% | 35 | 10 |
| | 75% | 21 | 1 | | 60% | 20 | 4 | | 45% | 23 | 7 |
| | 70% | 15 | 1 | | 55% | 14 | 3 | | 40% | 16 | 5 |
| | 65% | 11 | 1 | | 50% | 11 | 2 | | 35% | 12 | 4 |
| | 60% | 8 | 0 | | 45% | 8 | 2 | | 30% | 9 | 3 |
| | 55% | 7 | 0 | | 40% | 6 | 1 | | 25% | 7 | 2 |
| | 50% | 5 | 0 | | 35% | 5 | 1 | | | | |

†Level of significance = 5% and power = 80%

Generally, you use a table to look up the appropriate n and d^* values.

Επιλογή Μεθόδων Δειγματοληψίας

- Πληθυσμός που θα μελετηθεί
 - Μέγεθος και γεωγραφική κατανομή.
 - Ετερογένεια σε σχέση με τη μεταβλητή.
- Διαθέσιμοι κατάλογοι των μονάδων δειγματοληψίας.
- Επίπεδο ακρίβειας που απαιτείται.
- Διαθέσιμοι πόροι.
- Περιπτώσεις:
 - Μικρός πληθυσμός → απλή τυχαία
 - Μεγάλος πληθυσμός → συνδυασμός μεθόδων
 - Ετερογένεια μεταβλητής → στρωματοποιημένη
 - Γεωγραφικά διασκορπισμένος πληθυσμός → συνδυασμός μεθόδων

Ακρίβεια Δειγματοληψίας

Μέγεθος Δείγματος

| n=100 | | | n=1000 | | | n=10000 | | |
|--------|------|-----------|--------|------|-----------|---------|------|-----------|
| ΘΕΤΙΚΑ | % | 95% CI | ΘΕΤΙΚΑ | % | 95% CI | ΘΕΤΙΚΑ | % | 95% CI |
| 5 | 5,0 | 1,6-11,3 | 50 | 5,0 | 3,7-6,5 | 500 | 5,0 | 4,6-5,4 |
| 25 | 25,0 | 16,9-34,7 | 250 | 25,0 | 22,3-27,8 | 2500 | 25,0 | 24,2-25,9 |
| 50 | 50,0 | 39,8-60,2 | 500 | 50,0 | 46,9-53,1 | 5000 | 50,0 | 49,0-51,0 |
| 95 | 95,0 | 88,7-98,4 | 950 | 95,0 | 93,5-96,3 | 9500 | 95,0 | 94,6-95,4 |

Βήματα υπολογισμού του μεγέθους του δείγματος

- Εντοπισμός της μεταβλητής που θέλουμε να μελετήσουμε.
- Ποιος υπολογισμός πρέπει να εκτιμηθεί (% , μέση τιμή, λόγος κλπ.)
- Εκτιμήστε την αναμενόμενη συχνότητα του παράγοντα που μας ενδιαφέρει.
- Αποφασίστε για την επιθυμητή ακρίβεια της μέτρησης.
- Αποφασίστε για τον αποδεκτό κίνδυνο η μέτρηση να είναι εκτός των ορίων της πραγματικής τιμής στον πληθυσμό.
- Προσαρμόστε με βάση την επίδραση του σχεδιασμού.
- Προσαρμόστε με βάση το αναμενόμενο ποσοστό συμμετοχής.

Μαθηματικός τύπος υπολογισμού δείγματος

Simple random / systematic sampling

$$n = \frac{z^2 * p * q}{d^2} = \frac{1.96^2 * 0.15 * 0.85}{0.03^2} = 544$$

Cluster sampling

$$n = g * \frac{z^2 * p * q}{d^2} = \frac{2 * 1.96^2 * 0.15 * 0.85}{0.03^2} = 1088$$

z: alpha risk express in z-score

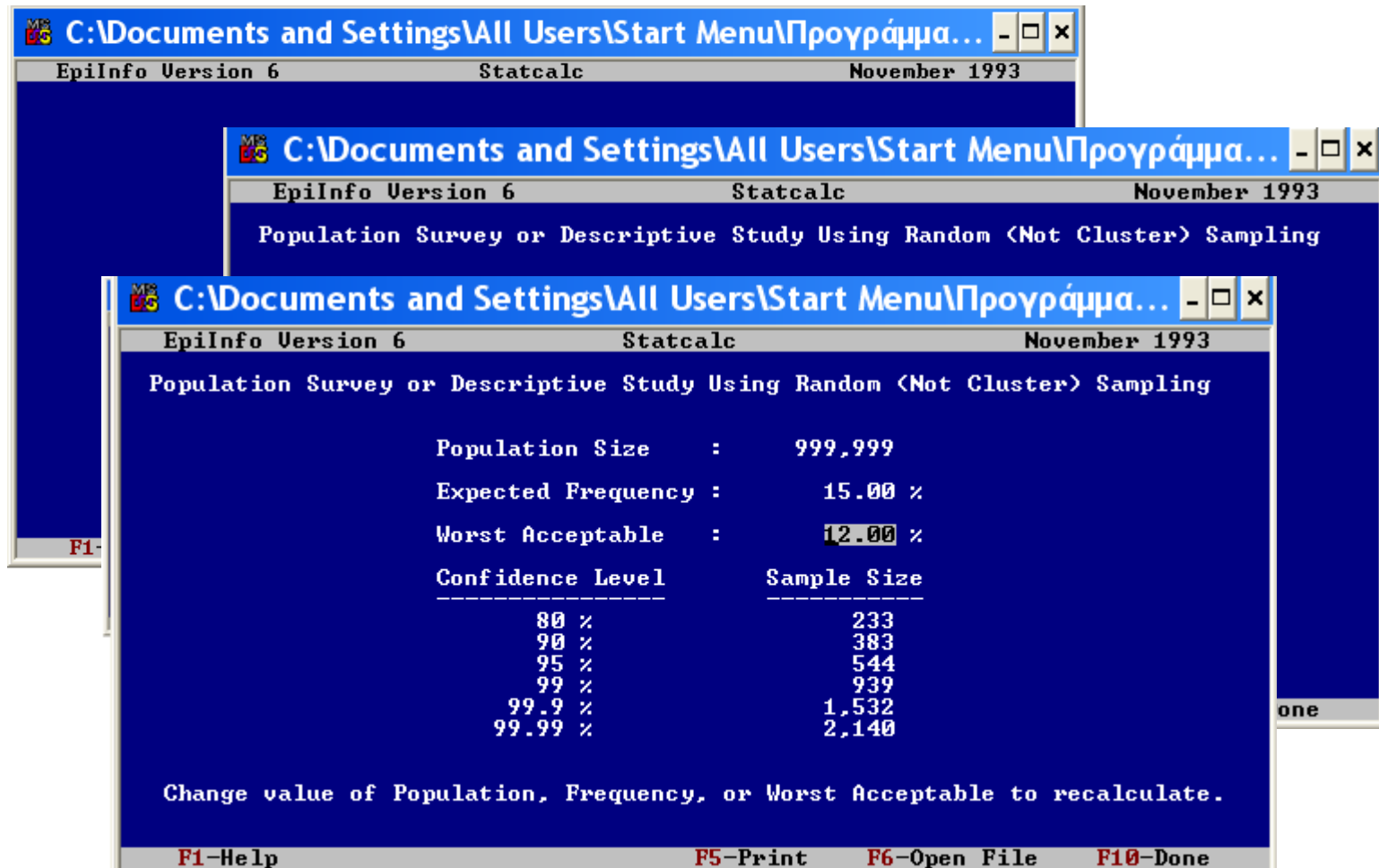
p: expected prevalence

q: 1 - p

d: absolute precision

g: design effect

EPI-INFO: υπολογισμός δείγματος



Θέση της δειγματοληψίας στις περιγραφικές επιδημιολογικές μελέτες

- Ορισμός στόχων
- Αναζήτηση πόρων
- Ορισμός του πληθυσμού υπό μελέτη
- Ορισμός των μεταβλητών-παραγόντων υπό μελέτη
- Σχεδιασμός της ανάλυσης (ερωτηματολόγιο)
- Δημιουργία πλάνου δειγματοληψίας
- Επιλέξτε το δείγμα
- Πιλοτική συλλογή δεδομένων
- Συλλογή δεδομένων
- Στατιστική ανάλυση δεδομένων
- Διάχυση αποτελεσμάτων
- Χρήση αποτελεσμάτων

Αποτελέσματα

- Τα δείγματα πιθανοτήτων είναι τα καλύτερα
- Να ληφθεί υπόψη ...
 - Άρνηση συμμετοχής
 - Απουσία
 - «Δεν ξέρω»