

## ΕΦΑΡΜΟΣΜΕΝΗ ΙΑΤΡΙΚΗ ΕΡΕΥΝΑ APPLIED MEDICAL RESEARCH

# Μεθοδολογία δειγματοληψίας στις επιδημιολογικές μελέτες

1. Εισαγωγή
2. Απλή τυχαία δειγματοληψία
3. Συστηματική τυχαία δειγματοληψία
4. Στρωματοποιημένη τυχαία δειγματοληψία
5. Ομαδοποιημένη δειγματοληψία
6. Μη τυχαία δειγματοληψία
7. Σύνοψη

### 1. ΕΙΣΑΓΩΓΗ

Η διεξαγωγή επιδημιολογικών μελετών βασίζεται σε δείγματα και όχι σε πληθυσμούς. Στην Επιδημιολογία, ο πληθυσμός (population) αποτελεί το σύνολο των δυνατών παρατηρήσεων που θα μπορούσαν θεωρητικά να πραγματοποιηθούν σε μια μελέτη, ενώ το δείγμα (sample) είναι ένα μέρος των παρατηρήσεων αυτών.<sup>1</sup> Για παράδειγμα, εάν σε μια μελέτη διερευνάται το μέσο βάρος των ενηλίκων Ελλήνων που κατοικούν στην Ελλάδα, τότε οι παρατηρήσεις είναι τα διάφορα βάρη και ο πληθυσμός είναι όλοι οι ενήλικες Έλληνες που κατοικούν στην Ελλάδα. Είναι σαφές ότι στη συγκεκριμένη περίπτωση είναι πρακτικά αδύνατον να διεξαχθεί μια μελέτη στην οποία θα μετρηθεί το βάρος όλων των ενηλίκων Ελλήνων που κατοικούν στην Ελλάδα. Αντίθετα, είναι εφικτή η εκπόνηση μιας μελέτης με συγκεκριμένο αριθμό ατόμων που συνιστούν το δείγμα της μελέτης. Επί πλέον, εάν σε μια μελέτη διερευνάται η μέση αρτηριακή πίεση ενός συγκεκριμένου πάσχοντα, τότε οι παρατηρήσεις είναι οι διάφορες μετρήσεις της αρτηριακής πίεσης του πάσχοντα και ο πληθυσμός είναι το σύνολο των δυνατών μετρήσεων της αρτηριακής πίεσης. Στην περίπτωση αυτή, είναι πρακτικά αδύνατον να διεξαχθεί μια μελέτη στην οποία θα μετρηθεί άπειρες φορές η αρτηριακή πίεση ενός πάσχοντα. Αντίθετα, είναι εφικτή η

διεξαγωγή μιας μελέτης με συγκεκριμένο αριθμό μετρήσεων της αρτηριακής πίεσης ενός πάσχοντα που συνιστούν το δείγμα της μελέτης.

Σημειώνεται ότι ο συνολικός αριθμός των παρατηρήσεων που αφορούν σε ένα συγκεκριμένο πληθυσμό μπορεί να είναι άπειρος ή πεπερασμένος. Στο παραπάνω παράδειγμα μέτρησης του μέσου βάρους των ενηλίκων Ελλήνων που κατοικούν στην Ελλάδα, ο πληθυσμός του συνολικού αριθμού των παρατηρήσεων είναι πεπερασμένος, ενώ στο παράδειγμα μέτρησης της μέσης αρτηριακής πίεσης ενός συγκεκριμένου πάσχοντα ο πληθυσμός του συνολικού αριθμού των παρατηρήσεων είναι άπειρος.

Έτσι, στις επιδημιολογικές μελέτες χρησιμοποιούνται δείγματα για την εξαγωγή συμπερασμάτων για τους αντίστοιχους πληθυσμούς. Χρησιμοποιώντας δείγματα, περιορίζεται βέβαια η ακρίβεια των μετρήσεων, ενώ στην πράξη σπάνια πραγματοποιούνται μελέτες με τη συμμετοχή ολόκληρων των μελετώμενων πληθυσμών. Η ακρίβεια πάντως των μετρήσεων στις επιδημιολογικές μελέτες μπορεί να ενισχυθεί με την αύξηση του αριθμού των παρατηρήσεων του δείγματος που λαμβάνεται από τους ερευνητές. Η αύξηση της ακρίβειας εξ άλλου επιτυγχάνεται και με τη μείωση της μεταβλητότητας των μετρήσεων, γεγονός όμως που δεν εξαρτάται από τους ερευνητές. Σκοπός των ερευνητών

ΑΡΧΕΙΑ ΕΛΛΗΝΙΚΗΣ ΙΑΤΡΙΚΗΣ 2012, 29(5):632-637  
ARCHIVES OF HELLENIC MEDICINE 2012, 29(5):632-637

Π. Γαλάνης

Εργαστήριο Οργάνωσης  
και Αξιολόγησης Υπηρεσιών Υγείας,  
Τμήμα Νοσηλευτικής, Εθνικό  
και Καποδιστριακό Πανεπιστήμιο  
Αθηνών, Αθήνα

Sampling methods in  
epidemiological studies

Abstract at the end of the article

### Λέξεις ευρητηρίου

Δείγμα  
Δειγματοληπτικό πλαίσιο  
Δειγματοληψία  
Στρωματοποιημένη δειγματοληψία  
Τυχαία δειγματοληψία

Υποβλήθηκε 14.6.2012  
Εγκρίθηκε 25.6.2012

είναι η πραγματοποίηση μελετών με όσο το δυνατόν λιγότερο κόστος και μικρότερο σφάλμα. Τα σφάλματα στις επιδημιολογικές μελέτες διακρίνονται στο τυχαίο και στο συστηματικό σφάλμα, με το πρώτο να αφορά στην ακρίβεια (reliability) και το δεύτερο στην εγκυρότητα (validity).<sup>2-4</sup> Έτσι, η μείωση του τυχαίου σφάλματος συνεπάγεται την αύξηση της ακρίβειας μιας μελέτης, ενώ η μείωση του συστηματικού σφάλματος συνεπάγεται την αύξηση της εγκυρότητας μιας μελέτης.

Οι ερευνητές σε μια επιδημιολογική μελέτη καλούνται να επιλέξουν ένα αντιπροσωπευτικό δείγμα από έναν πληθυσμό. Εάν το δείγμα μιας μελέτης είναι αντιπροσωπευτικό του πληθυσμού από τον οποίο προέρχεται, τότε είναι δυνατή η γενίκευση των αποτελεσμάτων της μελέτης στον πληθυσμό. Με τον τρόπο αυτόν επιτυγχάνεται η *εξωτερική εγκυρότητα* (external validity) μιας μελέτης.<sup>5</sup> Η επιλογή ενός μη αντιπροσωπευτικού δείγματος έχει ως αποτέλεσμα την απουσία εξωτερικής εγκυρότητας, οπότε δεν είναι δυνατή η γενίκευση των αποτελεσμάτων μιας μελέτης στον πληθυσμό από τον οποίο προέρχεται το δείγμα της μελέτης. Η επιλογή ενός δείγματος αντιπροσωπευτικού του πληθυσμού από τον οποίο προέρχεται, επιτυγχάνεται μέσω της τυχαίας επιλογής των παρατηρήσεων που συνιστούν το δείγμα. Για το λόγο αυτόν, στις επιδημιολογικές μελέτες είναι απαραίτητο η επιλογή του δείγματος να πραγματοποιείται με τυχαία δειγματοληψία (random sampling).<sup>6-14</sup>

Σημειώνεται ότι, σε ορισμένες περιπτώσεις, ο σχεδιασμός μιας μελέτης είναι τέτοιος που δεν είναι αναγκαία η διενέργεια τυχαίας δειγματοληψίας. Για παράδειγμα, στις μελέτες «ασθενών-μαρτύρων» συνήθως επιλέγονται όλοι οι ασθενείς του πληθυσμού-πηγή που εμφανίζουν τη μελετώμενη πάθηση σε ένα συγκεκριμένο χρονικό διάστημα, ενώ ο αριθμός των «μαρτύρων» συνήθως είναι πολύ μεγαλύτερος από τον αριθμό των ασθενών και γι' αυτό επιλέγεται ένα τυχαίο δείγμα «μαρτύρων».<sup>3</sup> Επί πλέον, στις κλινικές δοκιμές για την εκτίμηση της αποτελεσματικότητας των φαρμακευτικών αγωγών, η συμμετοχή των πασχόντων βασίζεται στα κριτήρια ένταξης (inclusion criteria) και στα κριτήρια αποκλεισμού (exclusion criteria) και όχι στην τυχαία επιλογή.<sup>3</sup> Οποσδήποτε, πάντως, η εφαρμογή της τυχαίας δειγματοληψίας είναι απαραίτητη στις συγχρονικές μελέτες και στις τυχαίοποιημένες ελεγχόμενες δοκιμές προληπτικής παρέμβασης για την εκτίμηση της αποτελεσματικότητας προληπτικών παρεμβάσεων, όπως π.χ. τα εμβόλια.

## 2. ΑΠΛΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Για την επιλογή ενός τυχαίου δείγματος, αρχικά, απαιτείται ο καθορισμός με ακρίβεια του πληθυσμού από τον οποίο

πρόκειται να προέλθει το εν λόγω δείγμα. Ο πληθυσμός αυτός είναι γνωστός ως *δειγματοληπτικό πλαίσιο* (sampling frame) και πρόκειται ουσιαστικά για την καταγραφή με τη μορφή καταλόγου όλων των παρατηρήσεων από τις οποίες πρόκειται να προέλθει το δείγμα μιας μελέτης.<sup>9,10</sup> Για παράδειγμα, για την επιλογή ενός τυχαίου δείγματος 10 φοιτητών από μια τάξη με 100 φοιτητές, το δειγματοληπτικό πλαίσιο συνίσταται από την καταγραφή των ονομάτων και των 100 φοιτητών. Ακολούθως, τα ονόματα των 100 φοιτητών καταγράφονται σε 100 διαφορετικά κομμάτια χαρτιού και τοποθετούνται σε ένα δοχείο, οπότε στη συνέχεια επιλέγονται με τυχαίο τρόπο 10 κομμάτια χαρτιού που αντιστοιχούν σε 10 φοιτητές. Με τον απλό αυτόν τρόπο και οι 100 φοιτητές έχουν την ίδια πιθανότητα να επιλεγούν στο δείγμα της μελέτης, οπότε το δείγμα που προκύπτει είναι τυχαίο. Επί πλέον, όλα τα πιθανά δείγματα με 10 φοιτητές έχουν την ίδια πιθανότητα να επιλεγούν, καθώς κάθε φοιτητής επιλέγεται ανεξάρτητα από τους υπόλοιπους. Η συγκεκριμένη μέθοδος δειγματοληψίας, στην οποία η πιθανότητα επιλογής μιας παρατήρησης είναι ίδια με την πιθανότητα επιλογής οποιασδήποτε άλλης παρατήρησης που περιλαμβάνεται στο δειγματοληπτικό πλαίσιο, είναι γνωστή ως *απλή τυχαία δειγματοληψία* (simple random sampling).<sup>6-14</sup>

Η συνήθης μέθοδος για την επιλογή τυχαίων αριθμών είναι η χρήση στατιστικών προγραμμάτων ανάλυσης δεδομένων, όπως το SPSS (Statistical Package for Social Sciences), το STATA κ.ά. Εξ άλλου, είναι δυνατή και η χρήση πινάκων τυχαίων αριθμών (tables of random numbers), που δημιουργούνται πλέον εύκολα σε διαδικτυακούς τόπους, όπως <http://stattrek.com/tables/random.aspx>, <http://graphpad.com/quickcalcs/randomN1.cfm> κ.ά. Στον πίνακα 1 παρουσιάζεται ένας πίνακας 100 τυχαίων αριθμών.

Για την επιλογή τυχαίων αριθμών από έναν πίνακα τυχαίων αριθμών εφαρμόζεται η ακόλουθη διαδικασία:

- Καθορίζεται επακριβώς το δειγματοληπτικό πλαίσιο του πληθυσμού από τον οποίο πρόκειται να προέλθει το δείγμα.
- Καθορίζεται το μέγεθος του δειγματοληπτικού πλαισίου, το οποίο συμβολίζεται με N.
- Καθορίζεται το μέγεθος του δείγματος, το οποίο συμβολίζεται με Δ.
- Καθορίζεται το εναρκτήριο σημείο επιλογής τυχαίων αριθμών στον πίνακα των τυχαίων αριθμών. Η επιλογή του εναρκτήριου σημείου πραγματοποιείται με κλειστά μάτια και την τοποθέτηση ενός δακτύλου στον πίνακα των τυχαίων αριθμών.
- Επιλέγεται η κατεύθυνση με την οποία θα γίνει η ανάγνωση του πίνακα (από πάνω προς τα κάτω, από

**Πίνακας 1.** Πίνακας 100 τυχαίων αριθμών.

Στήλη	Σειρά									
	1	2	3	4	5	6	7	8	9	10
1	031	028	038	009	096	068	019	092	022	029
2	035	018	067	082	051	043	021	061	042	024
3	071	037	012	086	005	033	098	003	011	062
4	039	034	057	052	085	089	017	045	087	090
5	007	026	016	074	078	015	004	084	063	060
6	070	041	027	100	023	054	050	099	013	083
7	048	053	093	047	056	002	069	044	010	065
8	094	066	001	049	077	095	058	008	020	036
9	046	059	073	055	030	014	080	079	088	075
10	000	097	040	025	076	032	081	006	091	072

αριστερά προς τα δεξιά ή από δεξιά προς τα αριστερά).

- Επιλέγονται τα πρώτα Δ ψηφία κατά την ανάγνωση του πίνακα, των οποίων τα τελευταία Α ψηφία βρίσκονται μεταξύ 0 και Ν. Εάν το Ν είναι διψήφιος αριθμός, τότε το Α=2, εάν το Ν είναι τριψήφιος αριθμός, τότε το Α=3, εάν το Ν είναι τετραψήφιος αριθμός, τότε το Α=4 κ.λπ.

Αναφορικά με το παραπάνω παράδειγμα, εφαρμόζοντας την απλή τυχαία δειγματοληψία για την επιλογή ενός τυχαίου δείγματος 10 φοιτητών από μια τάξη με 100 φοιτητές, για την επιλογή 10 τυχαίων αριθμών από έναν πίνακα 100 τυχαίων αριθμών (πίν. 1) εφαρμόζεται η ακόλουθη διαδικασία:

- Καθορίζεται επακριβώς το δειγματοληπτικό πλαίσιο του πληθυσμού, οπότε καταγράφονται με αριθμητική σειρά τα ονόματα και των 100 φοιτητών.
- Το μέγεθος του δειγματοληπτικού πλαισίου συμβολίζεται με Ν και είναι ίσο με 100.
- Το μέγεθος του δείγματος συμβολίζεται με Δ και είναι ίσο με 10.
- Ως εναρκτήριο σημείο επιλογής τυχαίων αριθμών στον πίνακα 1 επιλέγεται η σειρά 5 και η στήλη 1, δηλαδή ο αριθμός 096.
- Η ανάγνωση του πίνακα πραγματοποιείται από αριστερά προς τα δεξιά.
- Επιλέγονται τα πρώτα 10 (=Δ) ψηφία κατά την ανάγνωση του πίνακα, των οποίων τα τελευταία 3 (=Α) ψηφία βρίσκονται μεταξύ 0 και 100 (=Ν). Το Ν είναι τριψήφιος αριθμός, οπότε το Α=3.
- Το τυχαίο δείγμα των 10 φοιτητών με βάση τον πίνακα 1 είναι οι φοιτητές που αντιστοιχούν στους αριθμούς 96,

68, 19, 92, 22, 29, 35, 18, 67 και 82. Πρόκειται δηλαδή για τους αριθμούς που αντιστοιχούν στα σκιασμένα κελιά του πίνακα 1.

Η απλή τυχαία δειγματοληψία, μολονότι τεχνικά αποτελεί μια έγκυρη διαδικασία, είναι εξαιρετικά επίπονη και ιδιαίτερα όταν το μέγεθος του δείγματος είναι μεγάλο. Για το λόγο αυτόν, συνήθως εφαρμόζεται η *συστηματική τυχαία δειγματοληψία* (systematic random sampling), που είναι απλούστερη και ευκολότερη σε σχέση με την απλή τυχαία δειγματοληψία.<sup>6-14</sup>

### 3. ΣΥΣΤΗΜΑΤΙΚΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Η συστηματική τυχαία δειγματοληψία εφαρμόζεται συχνότερα σε σχέση με την απλή τυχαία δειγματοληψία. Στην περίπτωση της συστηματικής τυχαίας δειγματοληψίας εφαρμόζεται η ακόλουθη διαδικασία:

- Καθορίζεται επακριβώς το δειγματοληπτικό πλαίσιο του πληθυσμού από τον οποίο πρόκειται να προέλθει το δείγμα.
- Καθορίζεται το μέγεθος του δειγματοληπτικού πλαισίου, το οποίο συμβολίζεται με Ν.
- Καθορίζεται το μέγεθος του δείγματος, το οποίο συμβολίζεται με Δ.
- Διαιρείται το μέγεθος του δειγματοληπτικού πλαισίου με το μέγεθος του δείγματος, δηλαδή Ν/Δ, οπότε προκύπτει ένας αριθμός Α.
- Χρησιμοποιώντας έναν πίνακα τυχαίων αριθμών, επιλέγεται ένας τυχαίος αριθμός Χ μεταξύ 1 και Α. Ο τυχαίος αυτός αριθμός Χ αντιστοιχεί στην πρώτη παρατήρηση του δείγματος.
- Η δεύτερη παρατήρηση του δείγματος είναι η παρατήρηση Χ+Α του δειγματοληπτικού πλαισίου, η τρίτη παρατήρηση του δείγματος είναι η παρατήρηση Χ+2Α του δειγματοληπτικού πλαισίου, η τέταρτη παρατήρηση του δείγματος είναι η παρατήρηση Χ+3Α του δειγματοληπτικού πλαισίου κ.ο.κ.

Εφαρμόζοντας τη συστηματική τυχαία δειγματοληψία για την επιλογή ενός τυχαίου δείγματος 10 φοιτητών από μια τάξη με 100 φοιτητές εφαρμόζεται η ακόλουθη διαδικασία:

- Καθορίζεται επακριβώς το δειγματοληπτικό πλαίσιο του πληθυσμού, οπότε καταγράφονται με αριθμητική σειρά τα ονόματα και των 100 φοιτητών.
- Το μέγεθος του δειγματοληπτικού πλαισίου συμβολίζεται με Ν και είναι ίσο με 100.
- Το μέγεθος του δείγματος συμβολίζεται με Δ και είναι ίσο με 10.

- Διαιρείται το μέγεθος του δειγματοληπτικού πλαισίου με το μέγεθος του δείγματος, δηλαδή  $N/\Delta$ , οπότε προκύπτει 10 ( $=A$ ).
- Χρησιμοποιώντας έναν πίνακα τυχαίων αριθμών, επιλέγεται ένας τυχαίος αριθμός  $X$  μεταξύ 1 και 10. Ο τυχαίος αυτός αριθμός είναι το 5 και αντιστοιχεί στον πρώτο φοιτητή του δείγματος.
- Ο δεύτερος φοιτητής του δείγματος είναι ο 15ος φοιτητής ( $=5+10=15$ ) του δειγματοληπτικού πλαισίου, ο τρίτος φοιτητής του δείγματος είναι ο 25ος φοιτητής ( $=5+2\times 10=25$ ) του δειγματοληπτικού πλαισίου, ο τέταρτος φοιτητής του δείγματος είναι ο 35ος φοιτητής ( $5+3\times 10=35$ ) του δειγματοληπτικού πλαισίου κ.λπ. Έτσι, επιλέγονται τελικά 10 φοιτητές που συνιστούν το δείγμα της μελέτης.

#### 4. ΣΤΡΩΜΑΤΟΠΟΙΗΜΕΝΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Με τη στρωματοποιημένη τυχαία δειγματοληψία (stratified random sampling) εξασφαλίζεται η συμμετοχή συγκεκριμένων κατηγοριών ή, αλλιώς, στρωμάτων των παρατηρήσεων στο δείγμα της μελέτης.<sup>9,10</sup> Στην περίπτωση αυτή, ο πληθυσμός διαιρείται σε επί μέρους στρώματα, έτσι ώστε να δημιουργηθούν ομογενείς υποομάδες πριν από την έναρξη της δειγματοληψίας. Τα επί μέρους στρώματα που δημιουργούνται θα πρέπει να είναι αμοιβαίως αποκλειόμενα, δηλαδή κάθε παρατήρηση μπορεί να συμπεριληφθεί μόνο σε ένα στρώμα. Αρχικά, ο πληθυσμός διαιρείται σε επί μέρους στρώματα και έπειτα πραγματοποιείται απλή ή συστηματική τυχαία δειγματοληψία σε κάθε στρώμα ξεχωριστά. Με τον τρόπο αυτόν αυξάνεται η αντιπροσωπευτικότητα του δείγματος και μειώνεται το τυχαίο σφάλμα.

Η στρωματοποιημένη τυχαία δειγματοληψία είναι εξαιρετικά χρήσιμη στην περίπτωση κατά την οποία ορισμένες κατηγορίες παρατηρήσεων βρίσκονται σε μικρή αναλογία στον πληθυσμό. Για παράδειγμα, εάν το 5% ενός πληθυσμού ανήκει σε μια συγκεκριμένη εθνικότητα, τότε επιλέγοντας ένα δείγμα από τον πληθυσμό αυτόν είναι πιθανό να μη συμπεριλαμβάνονται άτομα της συγκεκριμένης εθνικότητας στο δείγμα της μελέτης. Σημειώνεται πάντως ότι όσο αυξάνεται το μέγεθος του δείγματος, τόσο μειώνεται η πιθανότητα να μη συμπεριλαμβάνονται συγκεκριμένες κατηγορίες παρατηρήσεων στο δείγμα της μελέτης. Σε κάθε περίπτωση όμως είναι προτιμότερο να εφαρμόζεται η στρωματοποιημένη τυχαία δειγματοληψία, έτσι ώστε να εξασφαλίζεται η ένταξη όλων των πιθανών κατηγοριών ενός πληθυσμού στο δείγμα της μελέτης.

Η στρωματοποιημένη τυχαία δειγματοληψία διακρίνεται

στην αναλογική (proportional) και στη μη αναλογική (non proportional). Στην αναλογική στρωματοποιημένη τυχαία δειγματοληψία, οι αναλογίες των στρωμάτων στο δείγμα είναι ίδιες με τις αναλογίες των στρωμάτων στον πληθυσμό από τον οποίο προέρχεται το δείγμα. Για παράδειγμα, εάν σε έναν πληθυσμό το 60% είναι γυναίκες και το 40% είναι άνδρες και το δείγμα της μελέτης αποτελείται από 1.000 άτομα, τότε θα δημιουργηθούν δύο στρώματα (γυναίκες και άνδρες), με το στρώμα των γυναικών να αποτελείται από 600 άτομα και το στρώμα των ανδρών να αποτελείται από 400 άτομα.

Στην περίπτωση που σκοπός μιας μελέτης είναι η πραγματοποίηση συγκρίσεων μεταξύ των διαφόρων κατηγοριών του δείγματος και ορισμένες κατηγορίες αποτελούνται από μικρό αριθμό παρατηρήσεων, τότε είναι πολύ πιθανό η μελέτη να χαρακτηρίζεται από περιορισμένη στατιστική ισχύ, με αποτέλεσμα να μην είναι εφικτή η διεξαγωγή συγκρίσεων με την εφαρμογή στατιστικών ελέγχων. Στη συγκεκριμένη περίπτωση είναι προτιμότερη η εφαρμογή της μη αναλογικής στρωματοποιημένης τυχαίας δειγματοληψίας έναντι της αναλογικής, έτσι ώστε να αυξηθεί ο αριθμός των παρατηρήσεων σε ορισμένες κατηγορίες και να είναι δυνατή η πραγματοποίηση συγκρίσεων. Για παράδειγμα, εάν σκοπός μιας μελέτης είναι η σύγκριση μεταξύ γυναικών που γέννησαν σε μαιευτήριο με γυναίκες που γέννησαν στο σπίτι, τότε θα πρέπει να διενεργηθεί μη αναλογική στρωματοποιημένη τυχαία δειγματοληψία, καθώς ο αριθμός των γυναικών που γεννούν στο σπίτι είναι εξαιρετικά μικρός.

Σε μια μελέτη, ο πληθυσμός των εργαζόμενων ιατρών και νοσηλευτών σε ένα νοσοκομείο αποτελείται από 200 άτομα και πιο συγκεκριμένα από 60 άνδρες ιατρούς, 20 γυναίκες ιατρούς, 20 νοσηλευτές και 100 νοσηλεύτριες. Έτσι, οι αναλογίες των εργαζόμενων στον πληθυσμό με βάση την επαγγελματική ιδιότητα και το φύλο είναι οι εξής:

- Άνδρες ιατροί= $60/200=0,3$  ή 30%
- Γυναίκες ιατροί= $20/200=0,1$  ή 10%
- Νοσηλευτές= $20/200=0,1$  ή 10%
- Νοσηλεύτριες= $100/200=0,5$  ή 50%

Πραγματοποιώντας αναλογική στρωματοποιημένη τυχαία δειγματοληψία για την επιλογή 50 εργαζόμενων, θα πρέπει:

- Το 30% των εργαζόμενων να είναι άνδρες ιατροί
- Το 10% των εργαζόμενων να είναι γυναίκες ιατροί
- Το 10% των εργαζόμενων να είναι νοσηλευτές
- Το 50% των εργαζόμενων να είναι νοσηλεύτριες.

Έτσι, στο δείγμα των 50 εργαζόμενων θα υπάρχουν:

- 15 άνδρες ιατροί ( $=0,3 \times 50$ )
- 5 γυναίκες ιατροί ( $=0,1 \times 50$ )
- 5 νοσηλευτές ( $=0,1 \times 50$ )
- 25 νοσηλεύτριες ( $=0,5 \times 50$ ).

Ακολούθως, η επιλογή των 50 εργαζόμενων του δείγματος από τις διάφορες κατηγορίες πραγματοποιείται με απλή ή συστηματική τυχαία δειγματοληψία.

## 5. ΟΜΑΔΟΠΟΙΗΜΕΝΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Η ομαδοποιημένη δειγματοληψία (cluster sampling) εφαρμόζεται στις περιπτώσεις όπου δεν χρειάζεται να καθοριστεί επακριβώς το δειγματοληπτικό πλαίσιο του πληθυσμού από τον οποίο πρόκειται να προέλθει το δείγμα.<sup>9,10</sup> Στην ομαδοποιημένη δειγματοληψία, αρχικά, ο πληθυσμός διαιρείται με τυχαίο τρόπο σε επί μέρους ομάδες (clusters), από τις οποίες στη συνέχεια προκύπτουν οι παρατηρήσεις του δείγματος. Για παράδειγμα, για την επιλογή ενός δείγματος μαθητών των δημοτικών σχολείων της Αττικής είναι εξαιρετικά επίπονο και χρονοβόρο να καταγραφούν όλοι οι μαθητές όλων των δημοτικών σχολείων της Αττικής. Στην περίπτωση αυτή πραγματοποιείται ομαδοποιημένη δειγματοληψία των δημοτικών σχολείων, οπότε αρχικά καταγράφονται επακριβώς τα δημοτικά σχολεία της Αττικής και έπειτα επιλέγεται ένα τυχαίο δείγμα από τα συγκεκριμένα σχολεία. Ακολούθως, όλοι οι μαθητές των δημοτικών σχολείων που επιλέχθηκαν αποτελούν το τυχαίο δείγμα της μελέτης. Είναι δυνατόν βέβαια να επιλεγεί ένα τυχαίο δείγμα των μαθητών των δημοτικών σχολείων που επιλέχθηκαν με την ομαδοποιημένη δειγματοληψία, χρησιμοποιώντας όμως πλέον την απλή ή τη συστηματική τυχαία δειγματοληψία.

Σημειώνεται ότι στην ομαδοποιημένη δειγματοληψία χρειάζεται ιδιαίτερη προσοχή στη διαίρεση του πληθυσμού σε επί μέρους ομάδες, έτσι ώστε οι ομάδες αυτές να είναι όσο το δυνατόν πιο αντιπροσωπευτικές του πληθυσμού από τον οποίο προέρχονται. Αναφορικά με το παραπάνω παράδειγμα με τους μαθητές των δημοτικών σχολείων της Αττικής, η επιλογή των επί μέρους ομάδων, δηλαδή των

σχολείων, θα πρέπει να πραγματοποιηθεί με τυχαίο τρόπο σύμφωνα με τη γεωγραφική κατανομή των σχολείων.

## 6. ΜΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Η μη τυχαία δειγματοληψία (non random sampling) δεν πρέπει να εφαρμόζεται στις επιδημιολογικές μελέτες, καθώς δεν συμπεριλαμβάνει την τυχαία επιλογή των παρατηρήσεων του δείγματος. Εφαρμόζοντας τη μη τυχαία δειγματοληψία προκύπτουν δείγματα ευκολίας (convenience samples), που πρέπει να αποτελούν την τελευταία επιλογή στο σχεδιασμό των επιδημιολογικών μελετών. Στην περίπτωση αυτή, δεν καθορίζεται επακριβώς το δειγματοληπτικό πλαίσιο του πληθυσμού από τον οποίο προέρχεται το δείγμα της μελέτης και απλά επιλέγονται ορισμένες παρατηρήσεις. Είναι σαφές ότι δεν εφαρμόζεται η τυχαία δειγματοληψία, οπότε είναι αδύνατη και η γενίκευση των συμπερασμάτων της μελέτης στον πληθυσμό από τον οποίο προέρχεται το δείγμα. Για παράδειγμα, ένα δείγμα χρηστών ναρκωτικών ουσιών που εθελοντικά συμμετείχαν σε μια μελέτη, έπειτα από σχετική διαφήμιση στα μέσα μαζικής ενημέρωσης, δεν αποτελεί τυχαίο δείγμα, καθώς δεν καταγράφηκε επακριβώς ο πληθυσμός των χρηστών ναρκωτικών ουσιών από τον οποίο θα μπορούσαν να προέλθουν οι παρατηρήσεις του δείγματος.

## 7. ΣΥΝΟΨΗ

Η εφαρμογή της τυχαίας δειγματοληψίας είναι απαραίτητη για το σχεδιασμό και την πραγματοποίηση επιδημιολογικών μελετών με ακριβή και έγκυρο τρόπο. Πρέπει να αποφεύγονται με κάθε τρόπο τα μη τυχαία δείγματα, έτσι ώστε τα συμπεράσματα των μελετών να είναι ασφαλή. Η γενίκευση εξ άλλου των συμπερασμάτων μιας μελέτης στον πληθυσμό από τον οποίο προήλθε το δείγμα της μελέτης μπορεί να πραγματοποιηθεί μόνο στην περίπτωση τυχαίας επιλογής του δείγματος. Έτσι, στη φάση του σχεδιασμού των επιδημιολογικών μελετών, η κατάλληλη επιλογή του δείγματος είναι πρωταρχικής σημασίας και χρειάζεται ιδιαίτερη προσοχή για την αποφυγή σφαλμάτων και παρερμηνειών.

## ABSTRACT

## Sampling methods in epidemiological studies

P. GALANIS

Center for Health Services Management and Evaluation, Department of Nursing, National and Kapodistrian University of Athens, Athens, Greece

Archives of Hellenic Medicine 2012, 29(5):632–637

Epidemiological studies rely on data from samples rather than populations. In epidemiology, “population” refers to all the possible observations that could theoretically be made in a study, while “sample” is a part of these observations. By using samples the precision of measurements is decreased, but in practice studies are rarely conducted with the participation of entire populations under study. Researchers conducting an epidemiological study have to choose a representative sample from a population. When the sample is representative of the population that it is derived from it is possible to generalize the results of the study of the sample to the whole population. In this way, external validity of the study is achieved. Selection of a representative sample of the population for epidemiological study is achieved by random sampling. The possible methods of random sampling are simple, systematic, stratified and cluster random sampling. In order to select a random sample, the first step is to define with precision the study population that gives rise to the sample. This population is known as the sampling frame and can be thought of as a list of all the observations from which the sample is to be derived. In random sampling, the probability of an observation being selected is equal to the probability of any observation of the sampling frame being selected. Participation of certain categories or strata of observations in a sample study is achieved by stratified random sampling, which increases the representativeness of the sample and decreases random error. Stratified random sampling can be separated into proportional and non proportional. In proportional stratified random sampling, the proportions of the strata in the sample are identical to the proportions of the strata in the population from which that sample is derived. Cluster sampling is applied in cases where the sampling frame of the population does not need to be determined with precision. In cluster sampling, the first step is to divide the population in a random way into separate clusters, following which the observations of the sample arise from the separate clusters.

**Key words:** Random sampling, Sample, Sampling, Sampling frame, Stratified sampling

## Βιβλιογραφία

- EVERITT BS. *Medical statistics from A to Z*. 2nd ed. Cambridge University, Press, Cambridge, 2006:179, 206
- ΓΑΛΑΝΗΣ ΠΑ, ΣΠΑΡΟΣ ΛΔ. *Εγχειρίδιο Επιδημιολογίας*. ΒΗΤΑ Ιατρικές Εκδόσεις, Αθήνα, 2010:129–140, 205–214
- ΓΑΛΑΝΗΣ ΠΑ, ΣΠΑΡΟΣ ΛΔ. *Κλινική και επιδημιολογική έρευνα. Βασικές έννοιες*. ΒΗΤΑ Ιατρικές Εκδόσεις, Αθήνα, 2012
- ΓΑΛΑΝΗΣ Π, ΣΠΑΡΟΣ Λ. Συστηματικά σφάλματα στις επιδημιολογικές μελέτες. *Αρχ Ελλ Ιατρ* 2007, 24:373–388
- FLETCHER RW, FLETCHER SW. *Clinical epidemiology. The essentials*. 4th ed. Lippincott Williams & Wilkins, Philadelphia, 2005:6–15
- RUMSEY D. *Statistics for dummies*. Wiley Publishing Inc, Indianapolis, 2003:41–42
- BOWERS D. *Medical statistics from scratch. An introduction for health professionals*. 2nd ed. John Willey & Sons, New Jersey, 2008:71–75
- CHERNICK M, FRIIS R. *Introductory biostatistics for the health sciences*. John Willey & Sons, New Jersey, 2003:22–45
- LOHR SL. *Sampling: Design and analysis*. 2nd ed. Brooks/Cole, Cengage Learning, Boston, 2010:25–101, 165–218
- BLAND M. *An introduction to medical statistics*. 3rd ed. Oxford University Press, Oxford, 2000:27–46
- BOSLAUGH S, WATTERS P. *Statistics in a nutshell*. O'Reilly Media Inc, Cambridge, 2008:133–137
- GOOD P, HARDIN J. *Common errors in statistics*. 2nd ed. John Willey & Sons, New Jersey, 2006:3–10
- UTTS J. *Seeing through statistics*. 2nd ed. Duxbury Press, Pacific Grove, 2005:48–68
- HAND J. *Statistics: A very short introduction*. Oxford University Press, Oxford, 2008:48–54

Corresponding author:

P. Galanis, 14 Dikis street, GR-157 73 Athens, Greece  
e-mail: pegalan@nurs.uoa.gr