

1^ο ΓΥΜΝΑΣΙΟ ΠΕΥΚΩΝ

Γ ΓΥΜΝΑΣΙΟΥ

ΜΑΘΗΜΑ : ΤΕΧΝΟΛΟΓΙΑ

ΕΚΠΑΙΔΕΥΤΙΚΟΣ :

Τρασανίδης Γεώργιος, διπλ. Ηλεκ/γος Μηχανικός Msc ΠΕ12 05

Πληθυσμός (population) και δείγμα (sample)

Οι έννοιες του πληθυσμού και του δείγματος είναι πρωταρχικές στη στατιστική.

- **Πληθυσμός (population)** είναι ένα συνήθως μεγάλο σύνολο ερευνητικών υποκειμένων- ατόμων, αντικειμένων, δοκιμίων-, του οποίου ορισμένα χαρακτηριστικά θέλουμε να ερευνήσουμε.

Σε μια έρευνα, μας ενδιαφέρει η τιμή ενός χαρακτηριστικού του πληθυσμού που μελετάμε. Αυτό το χαρακτηριστικό που αναφέρεται στον **πληθυσμό** λέγεται **παράμετρος (parameter)** και τις περισσότερες φορές έχει σταθερή τιμή που είναι άγνωστη στον ερευνητή. Οι παράμετροι συμβολίζονται με τα μικρά γράμματα του ελληνικού αλφαβήτου : πχ. ο μέσος όρος του πληθυσμού συμβολίζεται συνήθως με το **μ** και το ποσοστό του πληθυσμού που έχει ένα συγκεκριμένο χαρακτηριστικό με το **π**.

Τις περισσότερες φορές όμως, επειδή είναι πρακτικά και τεχνικά αδύνατο, εξαιρετικά χρονοβόρο και οικονομικά ασύμφορο να συμπεριλάβουμε όλο τον πληθυσμό στην έρευνα μας, αρκούμαστε στη μελέτη ενός μέρους του πληθυσμού, του **δείγματος (sample)**, το οποίο και τελικά μελετούμε.

- **Δείγμα (sample)** είναι το υποσύνολο, το μέρος του πληθυσμού, το οποίο και τελικά ερευνούμε με στόχο την εξαγωγή συμπερασμάτων που να αφορούν σε όλο τον πληθυσμό. Κάθε αριθμητική ποσότητα που υπολογίζεται στο δείγμα για να εκτιμηθεί μια άγνωστη παράμετρος του πληθυσμού λέγεται **στατιστικό στοιχείο (statistic)**. Πχ. ο μέσος όρος του δείγματος συμβολίζεται συνήθως με το $\bar{x}$.

Άρα η παράμετρος (parameter) αναφέρεται στο πληθυσμό, ενώ το στατιστικό στοιχείο (statistic) στο δείγμα.

Το στατιστικό στοιχείο εξυπηρετεί δυο σκοπούς :

- ✓ την εκτίμηση (estimation) της άγνωστης παραμέτρου και
- ✓ την περιγραφή του δείγματος.

Παράμετρος Μέτρο που περιγράφει τον **πληθυσμό**.

Στατιστικό Στοιχείο Μέτρο που περιγράφει το **δείγμα**

- Η τιμή του στατιστικού στοιχείου δεν παραμένει σταθερή, αλλά εξαρτάται από το δείγμα, ενώ η τιμή της παραμέτρου είναι σταθερή. Όταν βέβαια έχουμε επιλέξει το δείγμα και υπολογίζουμε το στατιστικό στοιχείο, έχουμε μια συγκεκριμένη τιμή με την οποία

εκτιμούμε την άγνωστη σε εμάς παράμετρο. Η εκτιμητική που είναι η μια βασική μέθοδος της επαγωγικής στατιστικής, -η άλλη είναι ο έλεγχος υποθέσεων-, επιβεβαιώνει την ακρίβεια της εκτίμησης μας αυτής. Η πιθανότητα του σφάλματος υπάρχει πάντα, ανεξάρτητα από το μέγεθος του δείγματος. Αν ρωτήσουμε δέκα τυχαία επιλεγμένες διαφορετικές ομάδες χιλίων ατόμων την ίδια ερώτηση, τα αποτελέσματα δε θα είναι τα ίδια. Η απόκλιση του στατιστικού στοιχείου από την παράμετρο, που παρατηρείται από δείγμα σε δείγμα, είναι το δειγματοληπτικό σφάλμα ή **στατιστικό σφάλμα** (*sampling error, margin of error*). Όσο όμως μεγαλύτερο είναι το τυχαίο δείγμα, τόσο μικρότερο είναι το στατιστικό σφάλμα.

- Οι επιστημονικές δειγματοληψίες αναφέρουν πάντοτε το στατιστικό σφάλμα, συνήθως μεταξύ 3% και 5%, όπως και το **επίπεδο εμπιστοσύνης** για την εκτίμηση τους (*confidence level*), συνήθως 95%. Αυτό σημαίνει ότι υπάρχει το ενδεχόμενο 5 φορές στις εκατό ,5%, τα αποτελέσματα της δειγματοληψίας να είναι λάθος, ενώ στις υπόλοιπες 95 φορές, 95%, το στατιστικό στοιχείο θα διαφέρει το πολύ 3%-5% από την πραγματική τιμή της παραμέτρου.
- Για να μπορούν τα αποτελέσματα της έρευνας που προκύπτουν από τη μελέτη του δείγματος να γενικευθούν σε όλο τον πληθυσμό και για να είναι αξιόπιστη η έρευνα θα πρέπει να τηρούνται κάποιες προϋποθέσεις.
- Η θεωρία δειγματοληψίας αποτελεί ένα μεγάλο κεφάλαιο της μεθοδολογίας της στατιστικής. Το **δείγμα** μας **δεν** πρέπει να είναι **περιστασιακό** ή **συμπωματικό** (*convenient sample*) ούτε **εθελοντικό** (*volunteer sample*), διότι αυτά τα δείγματα δεν είναι **αντιπροσωπευτικά**, άρα και **αμερόληπτα** (*unbiased*), οι άνθρωποι δύσκολα είναι απροκατάληπτοι και ανεπηρέαστοι.
- Η λύση για την αμερόληπτη επιλογή αντιπροσωπευτικού δείγματος είναι η εφαρμογή **τυχαίων** διαδικασιών, ώστε κάθε υποκείμενο το πληθυσμού να έχει a priori (εκ των προτέρων) γνωστή και μη μηδενική πιθανότητα να συμπεριληφθεί στο δείγμα.

Έτσι επιλέγονται **πιθανολογικές** ή **πιθανοκρατικές δειγματοληπτικές μέθοδοι** που χρησιμοποιούν τυχαίους μηχανισμούς για την επιλογή του δείγματος (*probability sampling methods*), οι κυριότερες μεταξύ αυτών είναι :

α. απλή τυχαία δειγματοληψία (*simple random sampling*)

Το δείγμα επιλέγεται με τέτοιο τρόπο, ώστε κάθε μέλος του πληθυσμού να έχει την ίδια και ανεξάρτητη πιθανότητα να επιλεγεί. Ανεξάρτητη πιθανότητα επιλογής σημαίνει ότι η επιλογή ενός μέλους δεν επηρεάζει την πιθανότητα επιλογής κάποιου άλλου. Προϋποτίθεται ότι έχουμε ένα πλήρη κατάλογο με όλα τα μέλη του πληθυσμού. Η επιλογή γίνεται συνήθως με πίνακες τυχαίων αριθμών, ή στον υπολογιστή με κατάλληλους αλγορίθμους τυχαίων αριθμών (*random number generators*).

β. 1- ανά-ν συστηματική δειγματοληψία (1-in-n systematic sampling)

Από μια λίστα των μελών του πληθυσμού επιλέγεται τυχαία ένα από τα πρώτα και μετά κάθε νιοστό από τη λίστα. Για παράδειγμα εάν θέλουμε ένα συστηματικό δείγμα 5 ατόμων από ένα πληθυσμό 100 ατόμων, επιλέγουμε τυχαία ένα αριθμό από το 1 έως το 20 ($=100:5$), έστω το 11, και από τη διαθέσιμη λίστα επιλέγουμε τα υποκείμενα με αριθμό 11,31,51,71 και 91.

γ. κατά στρώματα δειγματοληψία (stratified sampling).

Πρόκειται για απλή τυχαία δειγματοληψία από ξεχωριστές υπομονάδες του πληθυσμού. Από ένα απλό τυχαίο δείγμα 100 ατόμων παίρνουμε χωριστά ένα απλό τυχαίο δείγμα 50 ανδρών και ένα απλό τυχαίο δείγμα 50 γυναικών και τα ενώνουμε. Αν διατηρήσουμε και στο δείγμα την αναλογία που έχουν τα στρώματα στον πληθυσμό, τότε το δείγμα είναι αναλογικό (*proportional*). Προϋπόθεση είναι να γνωρίζουμε την ποσοστιαία αναλογία κάθε κατηγορίας στον πληθυσμό, ώστε να απαιτήσουμε να διατηρηθεί στο μείγμα. Σε ένα Λύκειο με 500 μαθητές φοιτούν 165 μαθητές στην Α Λυκείου, 170 στη Β Λυκείου και 165 στην Γ Λυκείου. Υπολογίζοντας σε ποσοστά % το 33% ($= 165/500$) των μαθητών φοιτούν στην Α Λυκείου, το 34% στη Β Λυκείου και το 33% στην Γ Λυκείου. Για να πάρουμε ένα αναλογικό δείγμα 100 μαθητών θα επιλέξουμε τυχαία 33 μαθητές από την Α, 34 από τη Β και 33 από τη Γ Λυκείου.

4. κατά δεσμίδες δειγματοληψία (cluster sampling).

Τα υποκείμενα της έρευνας χωρίζονται σε δεσμίδες ή συστάδες. Επιλέγουμε τυχαία μια συστάδα, της οποίας όλα τα μέλη αποτελούν το δείγμα. Επιλέγουμε αυτήν τη δειγματοληψία συνήθως όταν δεν έχουμε αναλυτικούς καταλόγους με όλα τα υποκείμενα του πληθυσμού. Ένα κατά δεσμίδες δείγμα μαθητών Λυκείου του Νομού Θεσσαλονίκης είναι όλοι οι μαθητές που φοιτούν σε ένα τυχαία επιλεγμένο Λύκειο του νομού.

5. πολυσταδιακή δειγματοληψία (multistage sampling)

Πρόκειται για συνδυασμό δυο ή περισσότερων από τις παραπάνω μεθόδους.

Εάν θέλουμε ένα δείγμα 72 μαθητών Λυκείου από την Περιφέρεια Κεντρικής Μακεδονίας, μπορούμε σε πρώτο στάδιο να χωρίσουμε τα Λύκεια των 8 νομών της Περιφέρειας σε στρώματα ανάλογα με την περιοχή που βρίσκονται –αστική, ημιαστική, αγροτική-, και να διαλέξουμε τυχαία ένα Λύκειο από κάθε στρώμα (σύνολο $3 \times 8 = 24$ Λύκεια). Σε δεύτερο στάδιο επιλέγουμε με απλή τυχαία δειγματοληψία 3 μαθητές από κάθε τάξη καθενός από τα 24 Λύκεια ($72 = 3 \times 24$).

Ερωτήσεις σε μια έρευνα

Πρώτη ερώτηση σε μια έρευνα είναι : **τι** θέλουμε να μελετήσουμε, το θέμα, το αντικείμενο της έρευνας μας. Από την απάντηση αυτή προκύπτουν η θεματική περιοχή της έρευνας, τα χαρακτηριστικά και οι **μεταβλητές** του

πληθυσμού που επιθυμούμε να ερευνήσουμε, όπως και το είδος των μεταβλητών.

Κατόπιν **ποιους** θέλουμε να μελετήσουμε –τα υποκείμενα της έρευνας μας ανθρώπους, αντικείμενα, δοκίμια. Από τα υποκείμενα της έρευνας καθορίζουμε τον **πληθυσμό** της έρευνας και το μέγεθος του.

Ακολουθώς **πως** θα τα μελετήσουμε, η **μέθοδος** της έρευνας μας, οι διαδικασίες συλλογής στοιχείων, τα όργανα μέτρησης, οι μονάδες μέτρησης των χαρακτηριστικών και οι κλίμακες μέτρησης τους.