

## ΣΤΑΤΙΣΤΙΚΗ

### Εισαγωγή

Πολλές φορές ακούμε από τα μέσα μαζικής ενημέρωσης ή διαβάζουμε στα διάφορα έντυπα για τα αποτελέσματα μιας έρευνας που βασίζεται σε στατιστικά δεδομένα. Αυτές οι πληροφορίες, που αναφέρονται σε συγκεκριμένη έρευνα, συνήθως δίνονται με τη μορφή πινάκων ή διαγραμμάτων. Εκτός από την παρουσίαση αυτών των πληροφοριών, πολλές φορές ακούμε ή διαβάζουμε μελέτες και αναλύσεις των παρατηρήσεων που αναφέρονται στην έρευνα.

Στατιστική είναι η εφαρμοσμένη επιστήμη που συγκεντρώνει και παρουσιάζει πληροφορίες, αλλά ταυτόχρονα μελετά και αναλύει τις παρατηρήσεις, που αναφέρονται στην οικονομία, στη μετεωρολογία, στην υγεία, στην κοινωνιολογία, στον αθλητισμό κτλ.

Η Στατιστική περιλαμβάνει τόσο τις μεθόδους συλλογής και επεξεργασίας στοιχείων όσο και τις μεθόδους ανάλυσης και μελέτης αυτών. Από τη μελέτη αυτή προκύπτουν οι σχέσεις που υπάρχουν στα διάφορα φαινόμενα και διατυπώνονται συμπεράσματα που είναι χρήσιμα για τη λήψη αποφάσεων.

Ως ορισμό της «Στατιστικής» θα δώσουμε τον συνηθέστερο και πλέον γνωστό ορισμό του R.A. Fisher (1890-1962), πατέρα της σύγχρονης Στατιστικής:

Στατιστική είναι ένα σύνολο αρχών και μεθοδολογιών για:

- τον σχεδιασμό της διαδικασίας συλλογής δεδομένων
- τη συνοπτική και αποτελεσματική παρουσίασή τους
- την ανάλυση και εξαγωγή αντίστοιχων συμπερασμάτων.

Ο κλάδος της Στατιστικής που ασχολείται με τον πρώτο στόχο λέγεται **σχεδιασμός πειραμάτων**, ενώ με τον δεύτερο ασχολείται η **περιγραφική στατιστική**, που αποτελεί και το αντικείμενο μελέτης μας στη συνέχεια. Τέλος, η **επαγωγική στατιστική ή στατιστική συμπερασματολογία** περιλαμβάνει τις μεθόδους με τις οποίες γίνεται η προσέγγιση των χαρακτηριστικών ενός μεγάλου συνόλου δεδομένων, με τη μελέτη των χαρακτηριστικών ενός μικρού υποσυνόλου των δεδομένων.

## ΕΝΟΤΗΤΑ 2.1 : ΠΛΗΘΥΣΜΟΣ - ΔΕΙΓΜΑ - ΜΕΤΑΒΛΗΤΕΣ

### Διερεύνηση

Δυο φοιτητές κοινωνιολογίας έχουν ως θέμα εργασίας να μελετήσουν κατά πόσον οι νέοι ηλικίας μεταξύ 12 και 20 ετών όλης της χώρας συζητούν στα μέσα κοι-

ωνικής δικτύωσης θέματα που έχουν να κάνουν με την προσωπική τους ζωή και αν το γεγονός αυτό επιφέρει αρνητικές επιπτώσεις στους ίδιους. Για να βοηθηθούν στην έρευνά τους, συγκέντρωσαν από το διαδίκτυο τα αποτελέσματα παρόμοιων ερευνών που πραγματοποιήθηκαν σε άλλες χώρες.



Σύμφωνα με τα στοιχεία μιας έρευνας που διεξήχθη πρόσφατα (πηγή ΑΠΕ-ΜΠΕ), τα κοινωνικά δίκτυα δημιουργούν επιπρόσθετο στρες στα νέα παιδιά της Βρετανίας. Για τη διεξαγωγή της έρευνας ερωτήθηκαν 5.000 νέοι/ες, ηλικίας μεταξύ 12 και 20 ετών, από όλη τη χώρα. Εκ

των ερωτηθέντων, το 40% δήλωσε ότι αισθάνεται άσχημα όταν κανείς από τους «φίλους» τους δεν κάνει «like» στις «selfies» τους, ενώ το 35% απάντησε ότι το επίπεδο της αυτοπεποίθησής τους συνδέεται άμεσα από τον αριθμό των «ακολουθών» που έχουν. Περισσότερες λεπτομέρειες για τα αποτελέσματα της συγκεκριμένης έρευνας φαίνονται στη διπλανή εικόνα.

- 1) Ποιους/ες εκπροσωπούν οι 5.000 νέοι/ες της έρευνας;
- 2) Ποια μπορεί να ήταν τα ερωτήματα και ποια τα χαρακτηριστικά ως προς τα οποία εξετάστηκαν οι ερωτηθέντες;

**Θέματα για συζήτηση**

Στη συνέχεια, οι δυο φοιτητές ετοίμασαν ένα αντίστοιχο ερωτηματολόγιο για τη δική τους έρευνα και το έστειλαν στους φίλους τους και τους συμφοιτητές τους για να το απαντήσουν.

Να γράψετε μερικές από τις ερωτήσεις που ενδεχομένως να συμπεριέλαβαν στο ερωτηματολόγιό τους οι δυο φοιτητές, οι οποίες να επιδέχονται άλλες ποσοτική και άλλες ποιοτική απάντηση.

**Ερωτήσεις**

- 1) Να σχολιάσετε τον τρόπο που εργάστηκαν οι φοιτητές για να συλλέξουν τα δεδομένα τους. Ήταν κατάλληλο το δείγμα που επέλεξαν ώστε να είναι αξιόπιστη η έρευνα;

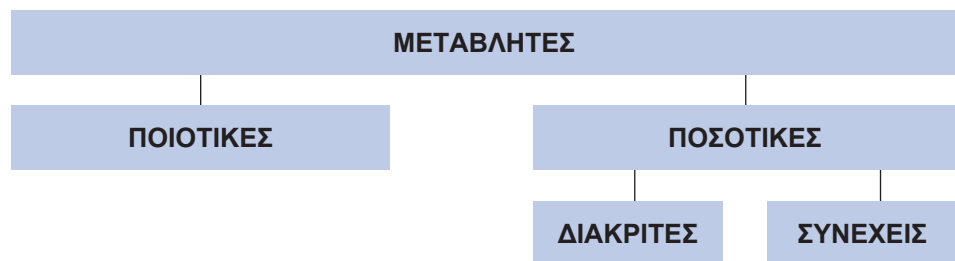
## Βασικές μαθηματικές έννοιες - Ιδέες - Διεργασίες

- Ορισμοί**
- Το σύνολο, στο οποίο επικεντρωνόμαστε σε μια έρευνα, ονομάζεται **πληθυσμός** της έρευνας. Τα στοιχεία του πληθυσμού συχνά αναφέρονται ως **άτομα** του πληθυσμού.
  - Τις περισσότερες φορές είναι πρακτικά πολύ δύσκολο ή και αδύνατο να προσεγγίσουμε καθένα από τα άτομα του πληθυσμού, για τεχνικούς λόγους ή λόγω χρόνου, κόστους κτλ. Στις περιπτώσεις αυτές επιλέγουμε ένα μέρος του πληθυσμού που είναι αντιπροσωπευτικό του πληθυσμού και το εξετάζουμε. Το μέρος αυτό του πληθυσμού λέγεται **δείγμα**.
  - Τα χαρακτηριστικά ως προς τα οποία εξετάζουμε τα άτομα ενός πληθυσμού λέγονται **μεταβλητές** και συμβολίζονται με ένα κεφαλαίο γράμμα X, Y, Z, ... κτλ.

Οι μεταβλητές διακρίνονται σε:

### Διάκριση μεταβλητών

- **Ποιοτικές** αν αναφέρονται σε ένα ποιοτικό χαρακτηριστικό του πληθυσμού, όπως είναι: τα χρώματα των αυτοκινήτων, το φύλο των μαθητών κτλ.
- **Ποσοτικές** αν αναφέρονται σε ένα ποσοτικό χαρακτηριστικό του πληθυσμού, όπως είναι: ο ετήσιος αριθμός των τροχαίων ατυχημάτων, το ύψος των μαθητών κτλ.
  - **Διακριτές** αν παίρνουν μεμονωμένες τιμές, όπως είναι: ο αριθμός των παιδιών των οικογενειών, το νούμερο των γυναικείων παπουτσιών ανά μισό πόντο κτλ.
  - **Συνεχείς** αν μπορούν να πάρουν οποιαδήποτε τιμή ενός διαστήματος, όπως είναι: το βάρος των μαθητών, ο χρόνος που χρειάζονται οι μαθητές για να απαντήσουν σε ένα διαγώνισμα κτλ.



## Εφαρμογές

### Εφαρμογή 1

Μια ομάδα μαθητών θέλει να προβεί σε πρόβλεψη για το ποιος θα είναι ο νέος πρόεδρος του μαθητικού συμβουλίου του σχολείου. Για τον λόγο αυτό τα μέλη της ομάδας κατέγραψαν την πρόθεση ψήφου των μαθητών ενός συγκεκριμένου τμήματος του σχολείου.

- 1) Ποιος είναι ο πληθυσμός της έρευνας;

- 2) Ποιο είναι το δείγμα;
- 3) Ποια είναι η μεταβλητή της έρευνας και ποιο το είδος της;
- 4) Να συζητήσετε πλεονεκτήματα και μειονεκτήματα της μεθόδου που επέλεξαν οι μαθητές, για να συγκεντρώσουν τις πληροφορίες και να κάνουν την πρόβλεψή τους.

### **Λύση**

- 1) Ο πληθυσμός της έρευνας είναι όλοι οι μαθητές του σχολείου.
- 2) Το δείγμα αποτελείται από τους μαθητές του τμήματος των οποίων καταγράφηκε η ψήφος τους.
- 3) Μεταβλητή της έρευνας είναι: «ο υποψήφιος που είναι προτιμώμενος για πρόεδρος του μαθητικού συμβουλίου». Η μεταβλητή αυτή είναι ποιοτική.
- 4) Το πλεονέκτημα της μεθόδου που επέλεξαν οι μαθητές είναι η ευκολία και η συντομία στο να πάρουν τις απαραίτητες απαντήσεις από τα άτομα του δείγματος. Το μειονέκτημα είναι ότι το δείγμα δεν είναι και αντιπροσωπευτικό του πληθυσμού, αφού αποτελείται από τους μαθητές ενός συγκεκριμένου τμήματος. Έτσι τα αποτελέσματα της έρευνας δε θα είναι αξιόπιστα, διότι οι υπόλοιποι μαθητές μπορεί να έχουν διαφορετική γνώμη.

### **Εφαρμογή 2**

Η τροχαία θέλει να κάνει μια έρευνα για τα τροχαία ατυχήματα που σημειώθηκαν το 2017 στον αυτοκινητόδρομο Αθηνών - Θεσσαλονίκης. Θέλει να εξετάσει τις πιο κάτω μεταβλητές:

- 1) τον τύπο του οχήματος που έχει εμπλακεί στο δυστύχημα,
- 2) την ταχύτητα του οχήματος,
- 3) τον αριθμό των τραυματιών,

Να χαρακτηρίσετε το είδος καθεμιάς από τις μεταβλητές.

### **Λύση**

Οι μεταβλητές χαρακτηρίζονται ως προς το είδος τους ως ακολούθως:

- 1) ποιοτική,
- 2) ποσοτική συνεχής,
- 3) ποσοτική διακριτή.

### **Ασκήσεις – Προβλήματα – Δραστηριότητες**

- 1) Ο δήμαρχος μιας πόλης θέλει να διερευνήσει τα προβλήματα που αντιμετωπίζει η πόλη του, ώστε να δώσει έμφαση στην επίλυση αυτών. Για τον λόγο αυτό, ανέθεσε σε μια εταιρεία δημοσκοπήσεων μια έρευνα κατά την οποία 500 δημότες κλήθηκαν να δηλώσουν ποιο πρόβλημα της πόλης θεωρούν το πιο σημαντικό. Στις παρακάτω ερωτήσεις να επιλέξετε τη σωστή απάντηση.

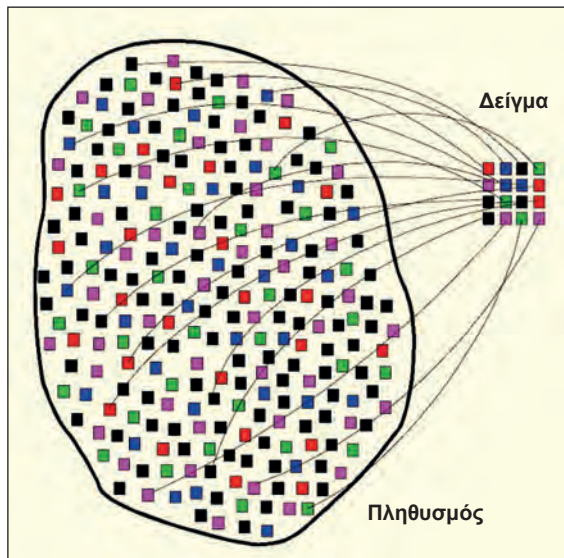
- 1) Ο πληθυσμός της έρευνας είναι:
    - 1) Οι 500 δημότες που ρωτήθηκαν.
    - 2) Όλοι οι δημότες της πόλης.
    - 3) Όλοι οι Έλληνες πολίτες.
  - 2) Το δείγμα αποτελούν:
    - 1) Οι 500 δημότες που ρωτήθηκαν.
    - 2) Όλοι οι δημότες της πόλης.
    - 3) Όλοι οι Έλληνες πολίτες.
  - 3) Η μεταβλητή της έρευνας είναι:
    - 1) Τα προβλήματα της πόλης.
    - 2) Τα προβλήματα των δημοτών.
    - 3) Το σημαντικότερο πρόβλημα της πόλης.
- 2)** Στις παρακάτω περιπτώσεις ποιες μπορεί να είναι οι μεταβλητές που μας ενδιαφέρουν; Να γίνει η διάκρισή τους σε ποιοτικές ή ποσοτικές και να αναφερθούν μερικές δυνατές τιμές τους:
- 1) Εξετάζουμε ένα δείγμα υπαλλήλων μιας εταιρείας.
  - 2) Εξετάζουμε ένα δείγμα προϊόντων από μια παραγωγή.
  - 3) Εξετάζουμε ένα δείγμα τηλεθεατών.
  - 4) Εξετάζουμε τους καλαθοσφαιριστές μιας ομάδας σε έναν αγώνα.
- 3)** Για τις ανάγκες μιας έρευνας συγκεντρώσαμε στοιχεία από διερχόμενα οχήματα, σε κάποιο συγκεκριμένο σημείο της πόλης, κατά τη διάρκεια ενός εικοσιτετραώρου. Στον επόμενο πίνακα παρουσιάζονται μερικά από τα στοιχεία αυτά.

Είδος οχήματος	Χρώμα οχήματος	Ταχύτητα σε Km/h	Πλήθος επιβατών
Φορηγό	Κόκκινο	26	2
Αυτοκίνητο ΙΧ	Γκρι	38	4
Ποδήλατο	Πράσινο	13	1
Αυτοκίνητο ΙΧ	Κόκκινο	52	2
Λεωφορείο	Λευκό	34	25
Αυτοκίνητο ΙΧ	Λευκό	45	1
Μοτοσικλέτα	Μαύρο	62	2

Ποιες είναι οι μεταβλητές της έρευνας και ποιο το είδος τους;

- 4)** Σε μια εκπομπή δημόσιας συζήτησης, συγκεκριμένου τηλεοπτικού καναλιού, το κοινό καλείται να ψηφίσει αν συμφωνεί με την Α άποψη ή τη Β άποψη. Να αναφέρετε δύο λόγους για τους οποίους τα αποτελέσματα της ψηφοφορίας δεν μπορεί να γενικευτούν σε ολόκληρο τον πληθυσμό της χώρας.

## Πρόσθετο υλικό



Η διαδικασία επιλογής ενός δείγματος από τον πληθυσμό ονομάζεται **δειγματοληψία**. Όπως είπαμε, προκειμένου να γενικεύσουμε τα συμπεράσματα που θα εξαγάγουμε από το δείγμα σε ολόκληρο τον πληθυσμό, έχει ιδιαίτερη σημασία το δείγμα να είναι αντιπροσωπευτικό του πληθυσμού. Αυτό σημαίνει ότι τα χαρακτηριστικά που θεωρούμε ότι επηρεάζουν τις μεταβλητές που μας ενδιαφέρουν

κατανέμονται με τον ίδιο τρόπο στο δείγμα και στον πληθυσμό. Στο παράδειγμα της Εφαρμογής 1, η σωστή πρακτική θα ήταν να συμπεριλάβουμε στο δείγμα μας μαθητές από κάθε τμήμα. Θα ήταν επίσης σωστό να συμπεριλάβουμε στο δείγμα μας αγόρια και κορίτσια με την ίδια περίπου αναλογία που βρίσκονται στον σχολικό πληθυσμό.

Ένας απλός τρόπος δειγματοληψίας που επιτυγχάνει στατιστικά την αντιπροσωπευτικότητα είναι η απλή τυχαία δειγματοληψία, στην οποία επιλέγουμε τυχαία τα μέλη του πληθυσμού που θα συμπεριληφθούν στο δείγμα. Για παράδειγμα, αν από έναν πληθυσμό 4.000.000 ανδρών και 4.000.000 γυναικών επιλέξουμε τυχαία 1.000 άτομα, τότε με πιθανότητα περίπου 99% το πλήθος των γυναικών που επιλέξαμε θα είναι μεταξύ 460 και 540.

Όσο μεγαλύτερο είναι το μέγεθος του δείγματος, τόσο πιο βέβαιοι μπορούμε να είμαστε ότι τα αποτελέσματα που θα πάρουμε από το δείγμα μας γενικεύονται στον πληθυσμό.

### Ιστορικά στοιχεία

Στις προεδρικές εκλογές στις ΗΠΑ το 1936 υπήρξε πλήρης αποτυχία της πρόγνωσης των εκλογικών αποτελεσμάτων από το περιοδικό «Literary Digest». Η έρευνα, που διενήργησε το περιοδικό, στηρίχτηκε σε αρχικό δείγμα 10.000.000 ατόμων του πληθυσμού από τηλεφωνικούς καταλόγους, τον κατάλογο των συνδρομητών του περιοδικού και καταλόγους ιδιοκτητών αυτοκινήτων. Στο ερωτηματολόγιο απάντησαν περίπου 2.400.000 άτομα και από την ανάλυση που πραγματοποιήθηκε φάνηκε ότι το εκλογικό σώμα θα έδινε μεγάλη πλειοψηφία στους Ρεπουμπλικάνους με υποψήφιο πρόεδρο τον Landon. Η μέρα των εκλογών όμως

επιφύλαξε μια δυσάρεστη έκπληξη στους στατιστικούς αναλυτές της δημοσκόπησης. Ο Δημοκρατικός υποψήφιος Roosevelt εκλέχτηκε με την ιστορική πλειοψηφία 60,8%, σε αντίθεση με το 42,9% που είχε προβλέψει η δημοσκόπηση του «Literary Digest».

Γιατί νομίζετε η έρευνα απέτυχε και μάλιστα σε τόσο μεγάλο βαθμό;