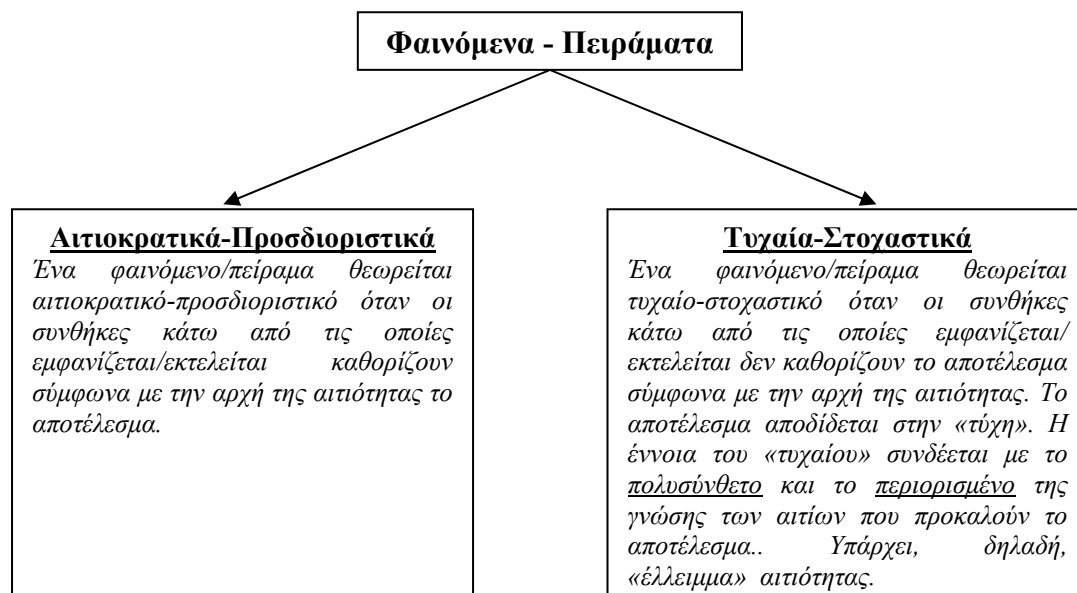


Στατιστική προσέγγιση προβλημάτων
Μια γενική επισκόπηση

Τα διάφορα προβλήματα (επιστημονικά ή άλλα) συνδέονται με φαινόμενα ή με πειράματα¹ τα οποία ταξινομούνται σε δύο γενικές κατηγορίες: στα **αιτιοκρατικά** ή **προσδιοριστικά/ντετερμινιστικά (deterministic)** και στα **τυχαία** ή **στοχαστικά (random, stochastic, non-deterministic)**.



Αν, για παράδειγμα, γνωρίζουμε το κεφάλαιο, το χρόνο και το επιτόκιο τότε γνωρίζουμε με βεβαιότητα και τον τόκο που πρέπει να εισπράξουμε στο συγκεκριμένο χρόνο ή αν γνωρίζουμε την κατανάλωση νερού και το κόστος ανά μονάδα κατανάλωσης τότε γνωρίζουμε με βεβαιότητα και το ποσό που πρέπει να πληρώσουμε. Όμως, μια ασφαλιστική εταιρεία δε γνωρίζει με βεβαιότητα ούτε τον αριθμό ούτε το ύψος των αποζημιώσεων που θα πληρώσει τον επόμενο μήνα. Επίσης, δε γνωρίζουμε με βεβαιότητα τον αριθμό των γεννήσεων που θα συμβούν την ερχόμενη εβδομάδα στην Κρήτη ή το αποτέλεσμα της θεραπείας ασθενών ηλικίας 30-40 ετών με ένα συγκεκριμένο φάρμακο ή την απόδοση μιας καλλιέργειας ή το ύψος των πωλήσεων μιας αυτοκινητοβιομηχανίας το επόμενο εξάμηνο.

Ένα αιτιοκρατικό-προσδιοριστικό φαινόμενο είναι δυνατόν να περιγραφεί με ένα μαθηματικό μοντέλο δηλαδή με ένα μαθηματικό ανάλογο/μίμηση του πραγματικού. Για παράδειγμα, η προσδιοριστική σχέση κεφάλαιο-επιτόκιο-χρόνος-τόκος μπορεί να περιγραφεί από ένα μαθηματικό τύπο ο οποίος μας επιτρέπει να προβλέψουμε με βεβαιότητα το αποτέλεσμα (τόκος) όταν γνωρίζουμε το κεφάλαιο, το επιτόκιο και το χρόνο. Όμως, ένα τυχαίο-στοχαστικό φαινόμενο όπως, για παράδειγμα, η σχέση τιμή-ζήτηση ενός προϊόντος δε μπορεί να περιγραφεί πλήρως με ένα μαθηματικό τύπο αφού η ζήτηση ενός προϊόντος οφείλεται, εκτός από την τιμή του, και σε άλλους παράγοντες-αιτίες όπως οι τιμές ομοειδών προϊόντων, το ύψος του εισοδήματος των καταναλωτών αλλά και σε άλλους που δε μπορούν να προσδιορισθούν με ακρίβεια. Επίσης, η σχέση απόδοση μιας καλλιέργειας-ποσότητα λιπάσματος είναι στοχαστική

¹ Ένα **πείραμα** διαφέρει από την παρατήρηση ενός **φαινομένου** κατά το ότι ο ερευνητής που εκτελεί το πείραμα παρεμβαίνει ενεργά επιβάλλοντας μια συγκεκριμένη μεταχείριση στα άτομα ή στα αντικείμενα, γενικότερα στα υποκείμενα, επί των οποίων αυτό εξελίσσεται. Αντίθετα, κατά την παρατήρηση ενός φαινομένου, μετράμε ή παρατηρούμε την κατάσταση των υποκειμένων επί των οποίων αυτό συμβαίνει χωρίς να προσπαθούμε να αλλάξουμε αυτή την κατάσταση με κάποια ειδική μεταχείριση.

και όχι προσδιοριστική αφού η απόδοση της καλλιέργειας δεν επηρεάζεται/εξαρτάται μόνο από την ποσότητα λιπάσματος αλλά και από άλλους παράγοντες όπως οι καιρικές συνθήκες, ο προσανατολισμός του αγρού, το έδαφος του αγρού, ο χρόνος σποράς αλλά και από άλλους που είτε δε γνωρίζουμε, είτε γνωρίζουμε αλλά δε μπορούμε να προσδιορίσουμε επακριβώς πώς επιδρούν στην απόδοση της καλλιέργειας. Συνεπώς, το «πολυσύνθετο» και το «περιορισμένο της γνώσης των αιτίων» που χαρακτηρίζουν τα τυχαία-στοχαστικά φαινόμενα και πειράματα απαιτούν επιστημονικά εργαλεία διαφορετικά από τα «συνήθη».

Ο κλάδος των *Μαθηματικών* που έχει ως αντικείμενο την έρευνα των νόμων που διέπουν τα τυχαία-στοχαστικά φαινόμενα και πειράματα ονομάζεται **Θεωρία Πιθανοτήτων**. Η σπουδαιότερη εφαρμογή της *Θεωρίας Πιθανοτήτων* είναι η ανάπτυξη **στατιστικών μεθόδων**. Οι στατιστικές μέθοδοι μας βοηθούν να βγάλουμε συμπεράσματα για όσα δε γνωρίζουμε ενώ η *Θεωρία Πιθανοτήτων* μας επιτρέπει να υπολογίσουμε πόσο βέβαιοι πρέπει να είμαστε για τα συμπεράσματά μας.

Σχόλιο 1.1: Σπάνια η πραγματικότητα μπορεί να περιγραφεί με προσδιοριστικά μαθηματικά μοντέλα. Όμως, πολλές φορές, προκειμένου πραγματικά προβλήματα να μελετηθούν στο πλαίσιο των «κλασικών» Μαθηματικών, γίνονται παραδοχές και απλουστεύσεις ώστε να «παρακάμπτεται» το τυχαίο και έτσι να είναι δυνατή η κατασκευή μαθηματικών προσδιοριστικών μοντέλων. Για παράδειγμα, η βολή πυροβόλου υπό ορισμένη γωνία και ορισμένη αρχική ταχύτητα, ενώ στο πλαίσιο των Μαθηματικών και της Φυσικής μελετάται ως αιτιοκρατικό πείραμα, στην πραγματικότητα είναι στοχαστικό αφού η τροχιά του βλήματος δεν επηρεάζεται μόνο από την αρχική ταχύτητα και τη γωνία βολής αλλά και από άλλους αστάθμητους παράγοντες. Η πραγματική τροχιά και το πραγματικό σημείο πτώσης, φυσικά, διαφέρουν από τα αντίστοιχα θεωρητικά. Επίσης, είναι φανερό ότι ερωτήματα όπως τα παρακάτω δε μπορούν να απαντηθούν από το προσδιοριστικό μαθηματικό μοντέλο που προκύπτει μετά από «απλούστευση» της πραγματικότητας μέσω παραδοχών.

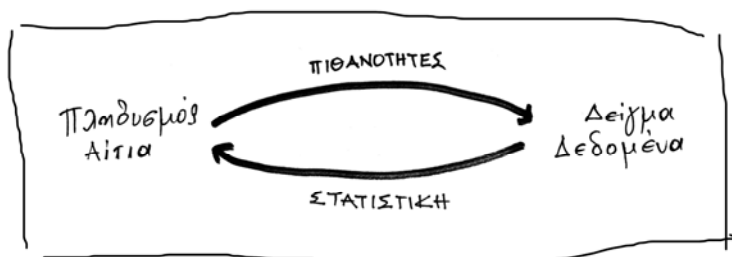
- ✓ Ποιο ποσοστό βλημάτων κατά μέσο όρο πλήττει το στόχο;
- ✓ Πόσα βλήματα πρέπει να χρησιμοποιηθούν για να πληγεί ο στόχος με ικανοποιητική ακρίβεια;
- ✓ Τι μέτρα πρέπει να ληφθούν για να περιορισθεί η διασπορά των σημείων πτώσης των βλημάτων;



Ένα φαινόμενο ή πείραμα δε **συμβαίνει/εκτελείται** αφηρημένα αλλά **σε ένα σύνολο υποκειμένων** (ατόμων, αντικειμένων, τόπων ή άλλων οντοτήτων) που έχουν ένα ή περισσότερα κοινά χαρακτηριστικά. Μετρώντας ή παρατηρώντας τα χαρακτηριστικά των υποκειμένων στα οποία συμβαίνει ή εκτελείται το φαινόμενο ή το πείραμα, δηλαδή, αποδίδοντας τιμές σε αυτά, παίρνουμε ένα σύνολο δεδομένων-παρατηρήσεων² από τα οποία προσπαθούμε να κατανοήσουμε τα αίτια και να εξηγήσουμε το φαινόμενο ή το πείραμα και να οδηγηθούμε στην κατασκευή ενός μοντέλου που θα το περιγράψει. Όμως, δεν είναι πάντοτε δυνατό να έχουμε στη διάθεσή μας όλες τις τιμές των χαρακτηριστικών του φαινομένου ή του πειράματος που μελετάμε. Συνήθως, μέρος μόνο των τιμών των χαρακτηριστικών έχουμε στη διάθεσή μας. Επίσης, λόγω του *πολυσύνθετου* της πραγματικότητας δεν έχουμε πλήρη εικόνα όλων των «χαρακτηριστικών». Στην αντιμετώπιση αυτών των δυσκολιών συνεισφέρουν η **Θεωρία Πιθανοτήτων** και η **Στατιστική**. Ειδικότερα, η διαδικασία

² **Δεδομένα/Παρατηρήσεις (data/observations):** στοιχεία που καταγράφονται κατά την παρατήρηση ενός φαινομένου ή κατά την εκτέλεση ενός πειράματος

μετάβασης από το «όλο» στο «μέρος» (από τον **πληθυσμό**³ στο **δείγμα**) και αντίστροφα, σε γενικές γραμμές, γίνεται ως εξής: *κάθε υποτιθέμενο μοντέλο για τον πληθυσμό συνεπάγεται, σύμφωνα με τη Θεωρία Πιθανοτήτων, ορισμένη συμπεριφορά για το δείγμα. Δηλαδή, η Θεωρία Πιθανοτήτων μας λέει τι πρέπει να περιμένουμε στο δείγμα, ή αλλιώς, τι περιμένουμε/προβλέπουμε για τα δεδομένα όταν μπορούμε να περιγράψουμε τα αίτια. Αντιστρόφως, μέσω της Στατιστικής, από τα δεδομένα συμπεραίνουμε για τα αίτια. Πιο συγκεκριμένα, η Στατιστική⁴ μας προσφέρει μεθόδους για να ελέγξουμε αν η συμπεριφορά του δείγματος διαφέρει σημαντικά από το αναμενόμενο με βάση τη Θεωρία Πιθανοτήτων. Αν αυτή η διαφορά, μεταξύ αναμενόμενου από τη θεωρία και παρατηρούμενου στο δείγμα, είναι σημαντική τότε πρέπει να αναζητηθεί άλλο μοντέλο για την περιγραφή του πληθυσμού.*



Προφανώς πρόκειται για μια διαδικασία ελέγχου της συμφωνίας (ή μη) **θεωρίας** και **εμπειρίας** όπου εφαρμόζεται τόσο η **παραγωγική** όσο και η **επαγωγική** αποδεικτική μέθοδος. Σε πρώτη φάση εφαρμόζεται η παραγωγική μέθοδος, δηλαδή, γίνεται μετάβαση από το *όλο* στο *μέρος*: υποθέτουμε κάποιο μοντέλο για τον *πληθυσμό* (το *όλο*) και μέσω της *Θεωρίας Πιθανοτήτων* συμπεραίνουμε για το *δείγμα* (το *μέρος*). Σε δεύτερη φάση εφαρμόζεται η επαγωγική μέθοδος, δηλαδή, γίνεται μετάβαση από το *μέρος* στο *όλο*: από τη συμπεριφορά του *δείγματος*, μέσω της *Στατιστικής*, συμπεραίνουμε για τον *πληθυσμό*. Δηλαδή, στη *στατιστική προσέγγιση* των προβλημάτων, συνυπάρχει ο *παραγωγικός* χαρακτήρας της *Μαθηματικής* επιστήμης με τον *επαγωγικό* χαρακτήρα των άλλων επιστημών. Το γεγονός αυτό εξηγεί την εφαρμοσιμότητα και τη χρησιμότητα αυτής της προσέγγισης σε ευρύτατο φάσμα επιστημών. Είναι επομένως φανερό ότι η *Στατιστική* μπορεί να συνεισφέρει σημαντικά σε οποιαδήποτε ερευνητική προσπάθεια που παράγει δεδομένα. Οι στατιστικές μέθοδοι προσφέρουν πολύτιμη υποστήριξη στις προσπάθειες που γίνονται για την ερμηνεία και την κατανόηση φαινομένων και καταστάσεων (*φυσικών, κοινωνικών, οικονομικών, πολιτικών, κ.ά.*), ανεξαρτήτως επιστημονικού/γνωστικού πεδίου. Κατ' επέκταση, η *Στατιστική* επηρεάζει τις

³**Πληθυσμός:** Στη Στατιστική με τον όρο **πληθυσμός (population)** ή **στατιστικός πληθυσμός (statistical population)** εννοούμε όλες τις τιμές που μπορεί να πάρει ένα κοινό χαρακτηριστικό μιας ομάδας υποκειμένων (ατόμων, αντικειμένων, τόπων και γενικότερα οποιωνδήποτε οντοτήτων) το οποίο μεταβάλλεται από υποκείμενο σε υποκείμενο (ή και στο ίδιο υποκείμενο) και ενδιαφερόμαστε να το μελετήσουμε. Κάθε υποκείμενο επί του οποίου μετράμε/παρατηρούμε το κοινό χαρακτηριστικό λέγεται **απλό στοιχείο** ή **δειγματοληπτική/πειραματική μονάδα (elementary unit/sampling unit)** και το κοινό χαρακτηριστικό τους **μεταβλητή (variable)**. Μάλιστα, όταν ένα κοινό χαρακτηριστικό παίρνει τιμές με βάση μια τυχαία διαδικασία, δηλαδή όταν η τιμή που κάθε φορά παίρνει καθορίζεται από το αποτέλεσμα ενός τυχαίου πειράματος ή φαινομένου, τότε ονομάζεται **τυχαία μεταβλητή (random variable)**. **Δείγμα (sample)** είναι ένα μέρος του πληθυσμού.

Σημειώνουμε ότι συχνά στη βιβλιογραφία ως δειγματοληπτική μονάδα ορίζεται ένα σύνολο απλών στοιχείων, δηλαδή, αυτές οι δύο έννοιες γενικά διακρίνονται. Στο πλαίσιο του παρόντος επιλέξαμε με τον όρο δειγματοληπτική μονάδα να εννοούμε όπως και με τον όρο απλό στοιχείο κάθε υποκείμενο επί του οποίου μετράμε/παρατηρούμε την τιμή μιας μεταβλητής.

⁴Με πιθανοθεωρητικούς συλλογισμούς και αξιοποιώντας αποτελέσματα της Θεωρίας Πιθανοτήτων!

διαδικασίες λήψης αποφάσεων και συνεπώς επηρεάζει ένα ευρύτατο φάσμα της ανθρώπινης δραστηριότητας. Ερευνητές, επαγγελματίες, επιχειρηματίες, πολιτικοί, κυβερνήσεις αλλά και κάθε πολίτης, συχνά αντιμετωπίζουν προβλήματα που σχετίζονται τόσο με τη συλλογή και την παρουσίαση δεδομένων όσο και με την ανάλυσή τους και την εξαγωγή **συμπερασμάτων** και **πληροφορίας** από αυτά ή απλά τους δημιουργούνται ερωτήματα σχετικά με την ερμηνεία και την κατανόηση αυτών των συμπερασμάτων και πληροφοριών. Δεδομένα και συμπεράσματα που συνάγονται από αυτά και σχετίζονται, για παράδειγμα, με την ανεργία, τις τιμές προϊόντων, τα ημερομίσθια, την παραγωγικότητα, τη γνώμη των πολιτών, τη συμπεριφορά των πολιτών, τις προτιμήσεις των καταναλωτών, την αποτελεσματικότητα μιας θεραπευτικής αγωγής, τη σεισμικότητα μιας περιοχής, έχουν άμεση ή έμμεση επίδραση στη ζωή μας.

Σημείωση 1.1: Η στατιστική προσέγγιση που περιγράψαμε σε γενικές γραμμές προηγουμένως, δεν είναι η μοναδική. Υπάρχουν και άλλες, όπως, η προσέγγιση της πιθανοφάνειας ή η Μπεϋζιανή προσέγγιση στις οποίες δε θα αναφερθούμε. Θα σημειώσουμε μόνο ότι, σχετικά πρόσφατα, έχουν αναπτυχθεί τεχνικές **διερευνητικής ανάλυσης δεδομένων (exploratory data analysis)** αφετηρία των οποίων δεν είναι ο πληθυσμός αλλά τα δεδομένα. Δηλαδή, δεν αναζητούμε απαντήσεις σε ερωτήματα που θέτουμε εκ των προτέρων (μοντέλα που υποθέτουμε ότι ισχύουν στον πληθυσμό) αλλά αφήνουμε τα δεδομένα “να μιλήσουν”. Αυτές οι τεχνικές είναι ιδιαιτέρως χρήσιμες όταν τα δεδομένα προϋπάρχουν των ερωτημάτων, όπως για παράδειγμα, δεδομένα από κρατικά αρχεία. Όμως, οι τεχνικές της διερευνητικής ανάλυσης δεδομένων είναι πολύ χρήσιμες και απαραίτητες ακόμη και στις πιο καλοσχεδιασμένες έρευνες γιατί μπορεί να αποκαλύψουν λάθη ή μια σημαντική επίδραση που δεν αναμενόταν ή να αναδείξουν ερωτήματα που δεν είχαμε σκεφθεί. ■

Η **Στατιστική** είναι η επιστήμη των δεδομένων. Μπορεί να ορισθεί ως «η επιστήμη που έχει αντικείμενο την ανάπτυξη μεθόδων για τη συλλογή, παρουσίαση, ανάλυση και ερμηνεία δεδομένων» (R.A.Fisher).

Οι στατιστικές μέθοδοι, με κριτήριο τους στόχους που εξυπηρετούν, ταξινομούνται σε τρεις γενικές κατηγορίες:

1. Μέθοδοι για το σχεδιασμό της διαδικασίας συλλογής των δεδομένων
2. Μέθοδοι για τη συνοπτική και την αποτελεσματική παρουσίαση των δεδομένων
3. Μέθοδοι για την εξαγωγή συμπερασμάτων

Με βάση αυτή την ταξινόμηση, η Στατιστική διαρθρώνεται σε τρεις κλάδους:

1. Σχεδιασμός Πειραμάτων (Experimental Design) και Θεωρία Δειγματοληψίας (Sampling Theory)

Είναι κλάδοι της Στατιστικής που, σε γενικές γραμμές, έχουν ως αντικείμενο την ανάπτυξη μεθόδων για τη συλλογή δεδομένων μέσω της εκτέλεσης πειραμάτων ή μέσω δειγματοληψιών αντίστοιχα.

2. Περιγραφική Στατιστική (Descriptive Statistics)

Είναι ο κλάδος της Στατιστικής που έχει ως αντικείμενο την ανάπτυξη μεθόδων για τη συνοπτική και εύληπτη παρουσίαση δεδομένων που προέρχονται από κάποιον πληθυσμό που μελετάμε.

3. Στατιστική Συμπερασματολογία (Statistical Inference)

Είναι ο κλάδος της Στατιστικής που έχει ως αντικείμενο την ανάπτυξη μεθόδων για την ανάλυση των δεδομένων και την εξαγωγή συμπερασμάτων για τον πληθυσμό από τον οποίο προέρχονται.

Πληθυσμός, Δείγμα και Στατιστική συμπερασματολογία

Για την εφαρμογή μεθόδων της στατιστικής συμπερασματολογίας ο πληθυσμός πρέπει να είναι **καλά ορισμένος**. Πρέπει να μπορούμε να αποφανθούμε με σαφήνεια επί ποίων μονάδων μετράμε/παρατηρούμε το χαρακτηριστικό που μελετάμε. Για παράδειγμα, αν ενδιαφερόμαστε να μελετήσουμε την επίδοση των μαθητών Γυμνασίου, ο πληθυσμός είναι καλά ορισμένος μόνο αν μπορούμε να απαντήσουμε με σαφήνεια (μονοσήμαντα) σε ερωτήσεις όπως: *μας ενδιαφέρει η μέτρηση της επίδοσης των μαθητών των ιδιωτικών Γυμνασίων; της Β' τάξης Γυμνασίου; νυχτερινών Γυμνασίων; ειδικών σχολείων; κτλ.* Επίσης, αν ενδιαφερόμαστε να μελετήσουμε το ποσοστό ανεργίας στη χώρα, πρέπει να ορισθεί με απόλυτη σαφήνεια ποιοι πολίτες θεωρούνται άνεργοι (*ποιες ομάδες ηλικιών περιλαμβάνονται; οι αλλοδαποί περιλαμβάνονται; οι φυλακισμένοι; ποιο είναι το χρονικό διάστημα αναφοράς; κτλ.*). Η κλινική δοκιμή ενός φαρμάκου για τη θεραπεία της ακμής πρέπει επίσης να αφορά ένα καλά ορισμένο πληθυσμό (*όλα τα άτομα που έχουν κάποιο τύπο ακμής; θα περιορίζεται σε όσα άτομα έχουν ένα συγκεκριμένο τύπο ακμής; ποιας ηλικίας; κτλ.*).

Οι πληθυσμοί διακρίνονται σε **πεπερασμένους** και σε **άπειρους**. Με κριτήριο το πλήθος των μονάδων επί των οποίων μετράμε/παρατηρούμε το κοινό χαρακτηριστικό που μελετάμε, φυσικά, άπειροι πληθυσμοί δεν υπάρχουν. Όμως, επειδή ως πληθυσμό θεωρούμε **όλες τις δυνατές τιμές** της υπό μελέτη μεταβλητής και δεδομένου ότι η παρατήρηση ενός φαινομένου ή ένα πείραμα μπορεί, θεωρητικά τουλάχιστον, να επαναληφθεί άπειρες φορές, τότε η έννοια «άπειρος πληθυσμός» έχει νόημα.

Ένας τρόπος για να μελετήσουμε ένα φαινόμενο είναι να εξετάσουμε όλες τις μονάδες επί των οποίων συμβαίνει. Αυτή η μέθοδος συλλογής δεδομένων, η εξέταση δηλαδή όλων των μονάδων επί των οποίων συμβαίνει το φαινόμενο που μελετάμε, ονομάζεται **ολική απογραφή** ή **απογραφή (census)**. Για παράδειγμα, η *Ελληνική Στατιστική Αρχή* κάνει κάθε δέκα χρόνια απογραφή του πληθυσμού ως προς διάφορα χαρακτηριστικά. Όμως, σε πολλές περιπτώσεις, η απογραφή είναι δύσκολη ή οικονομικά ασύμφορη ή και αδύνατη. Σε αυτές τις περιπτώσεις, όπως ήδη έχουμε αναφέρει, ο ερευνητής εξετάζει ένα μέρος των μονάδων που μελετάει, δηλαδή, παίρνει ένα **δείγμα** τιμών/μετρήσεων/παρατηρήσεων, και στη συνέχεια γενικεύει τα συμπεράσματά του για ολόκληρο τον πληθυσμό. Πώς μπορούμε όμως να επιλέξουμε ένα δείγμα **αντιπροσωπευτικό** του πληθυσμού από τον οποίο προέρχεται; Η απάντηση σε αυτό το ερώτημα δεν είναι εύκολη ούτε μπορεί να είναι απόλυτη. Απαντήσεις σε ερωτήματα σχετικά με τη διαδικασία επιλογής του δείγματος και με το μέγεθος του δείγματος δίνει η **Θεωρία Δειγματοληψίας**. Στο πλαίσιο του παρόντος θα αναφερθούμε μόνο στο νόημα της **τυχαίας δειγματοληψίας (random sampling)** και της **απλής τυχαίας δειγματοληψίας (simple random sampling)**.

Όλοι έχουμε δοκιμάσει μια κουταλιά σούπα για να κρίνουμε ως προς τη γεύση της όλη τη σούπα που βρίσκεται σε ένα πιάτο. Η σούπα όμως είναι ομοιογενής. Η γεύση μιας κουταλιάς αντιπροσωπεύει όλη την υπόλοιπη. Αλλά όταν πρόκειται για τη ζήτηση ενός προϊόντος, για τον αριθμό των ατυχημάτων, για το φόβο που αισθάνονται οι πολίτες απέναντι στο έγκλημα, για την επίδραση ενός φαρμάκου, για την απόδοση μιας καλλιέργειας ή για τη φορά του ίχνους της κίνησης των πάγων, υπάρχουν πολλοί παράγοντες που επιδρούν στο αποτέλεσμα που ποικίλουν από άτομο σε άτομο ή από

περιοχή σε περιοχή (γενικά, από υποκείμενο σε υποκείμενο). Η αναζήτηση και η κατανόηση των αιτίων αυτής της *διαφορετικότητας/μεταβλητότητας* είναι η κεντρική λειτουργία και η ουσία της *Στατιστικής*. **Η Στατιστική δεν θα είχε λόγο ύπαρξης αν δεν υπήρχε ποικιλομορφία/διαφορετικότητα/μεταβλητότητα. Σε έναν κόσμο ... «σούπα» με πανομοιότυπα χαρακτηριστικά και συμπεριφορές θα αρκούσε η μελέτη ενός μόνο υποκειμένου.**

Ένα κριτήριο επιλογής δείγματος θα μπορούσε να είναι η ευκολία. Για παράδειγμα, απευθυνόμαστε σε άτομα που βρίσκονται κοντά μας ή σε άτομα που περνούν από ένα πολυσύχναστο σημείο. Άλλα κριτήρια θα μπορούσαν να είναι η εθελοντική ανταπόκριση, δηλαδή, αυτοί που συμμετέχουν να επιλέγουν τους εαυτούς τους, η φιλικότητα, η ασφάλεια κτλ. Τέτοια όμως δείγματα προκαλούν **μεροληψίες** δηλαδή, **συστηματικές** διαφορές μεταξύ των αποτελεσμάτων που παίρνουμε από το δείγμα και της πραγματικότητας στον *πληθυσμό* διότι σε τέτοια δείγματα κατά κανόνα περιλαμβάνονται μετρήσεις/παρατηρήσεις από άτομα συγκεκριμένων κατηγοριών και αποκλείονται μετρήσεις/παρατηρήσεις από άτομα όλων των άλλων. Για παράδειγμα, μπορεί να αισθανόμαστε πιο ασφαλείς να πλησιάσουμε άτομα που δείχνουν ευγενικά και φιλικά αλλά έτσι οδηγούμαστε στο να αποκλείσουμε από το δείγμα μετρήσεις/παρατηρήσεις από άτομα μη φιλικά. Επίσης, σε μια τηλεφωνική έρευνα που πραγματοποιείται κατά τη διάρκεια μιας τηλεοπτικής εκπομπής, κατά κανόνα, ανταποκρίνονται άτομα με έντονα συναισθήματα, συνήθως αρνητικά. Έτσι, όταν ο εκφωνητής ρωτά τους τηλεθεατές αν *φοβούνται να βγουν έξω τη νύχτα λόγω της εγκληματικότητας*, κατά κανόνα θα τηλεφωνήσουν αυτοί που είναι εξαγριωμένοι με το έγκλημα παρά αυτοί που είναι πιο ψύχραιμοι.

Μεροληψίες: Είναι σφάλματα προς την ίδια κατεύθυνση γι' αυτό ονομάζονται και **συστηματικά σφάλματα (systematic errors)**. Μαζί με τα **μη συστηματικά/τυχαία σφάλματα (random errors)** αποτελούν τα **μη δειγματοληπτικά σφάλματα (non-sampling errors)**. Για τα μη δειγματοληπτικά σφάλματα δεν υπάρχει κάποιος γενικός κανόνας εκτίμησης του μεγέθους τους. Πηγές των μη δειγματοληπτικών σφαλμάτων είναι μεταξύ άλλων η λάθος επιλογή δείγματος, τα λάθη στα ερωτηματολόγια, τα λάθη συνέντευξης, τα λάθη επεξεργασίας κ.ά. Από αυτά, τα μη συστηματικά αλληλοαναιρούνται (ιδιαίτερα στα μεγάλα δείγματα).

Η προσωπική επιλογή είναι μια συνηθισμένη αιτία *μεροληψίας*. Για να μειωθεί επομένως η *μεροληψία* πρέπει να μειώσουμε την προσωπική επιλογή. Η απάντηση της *Θεωρίας Δειγματοληψίας* σε αυτό το πρόβλημα είναι η ανάπτυξη μεθόδων επιλογής **τυχαίων δειγμάτων**. Οι μέθοδοι αυτές ονομάζονται **σχέδια τυχαίας δειγματοληψίας**.

Υπάρχουν πολλά σχέδια *τυχαίας δειγματοληψίας*. Το πιο απλό είναι αυτό σύμφωνα με το οποίο ένα *δείγμα μεγέθους n* λαμβάνεται με τέτοιο τρόπο ώστε κάθε δυνατό *δείγμα μεγέθους n* να έχει την ίδια πιθανότητα να επιλεγεί. Ένα δείγμα που επιλέγεται με τέτοιο τρόπο ονομάζεται **απλό τυχαίο δείγμα (simple random sample)**.

Η τυχαία δειγματοληψία μπορεί να εξαλείψει τη μεροληψία που οφείλεται στη λάθος επιλογή δείγματος αλλά δεν εξαλείφει τη μεταβλητότητα. Η μεταβλητότητα από δείγμα σε δείγμα σε επαναλαμβανόμενα τυχαία δείγματα είναι αναπόφευκτη συνέπεια της μεταβλητότητας που υπάρχει στον πληθυσμό από τον οποίο αυτά προέρχονται.

Πώς μπορούμε επομένως να εμπιστευθούμε τα αποτελέσματα ενός *τυχαίου δείγματος*, ξέροντας ότι ένα δεύτερο *τυχαίο δείγμα* από τον ίδιο *πληθυσμό* θα έδινε διαφορετικό αποτέλεσμα; Πώς μπορούμε να βασίσουμε οικονομικές και πολιτικές αποφάσεις για την ανεργία, στο ποσοστό ανεργίας που δίνει ένα *τυχαίο δείγμα*, ξέροντας ότι το ποσοστό αυτό θα διαφέρει αν η δειγματοληψία επαναληφθεί;

Και όμως, στην πραγματικότητα μπορούμε να εμπιστευθούμε τα *τυχαία δείγματα*. Στο πλαίσιο της *Θεωρίας Δειγματοληψίας* έχει αναπτυχθεί μια μεγάλη ποικιλία σχεδίων τυχαίας δειγματοληψίας τα οποία καλύπτουν ένα ευρύ φάσμα περιπτώσεων. Από τα πιο γνωστά είναι η *στρωματοποιημένη δειγματοληψία (stratified sampling)*, η *συστηματική (systematic sampling)*, η *κατά συστάδες (cluster sampling)* και η *δισταδιακή κατά συστάδες δειγματοληψία (two stages cluster sampling)*. Όλα έχουν πλεονεκτήματα αλλά και μειονεκτήματα. Όμως, όλα έχουν ως κοινή βασική αρχή την *απλή τυχαία δειγματοληψία* γεγονός που εξασφαλίζει την πιθανοθεωρητική⁵ εκτίμηση του μεγέθους των **δειγματοληπτικών σφαλμάτων** αλλά και τον περιορισμό τους.

Δειγματοληπτικά σφάλματα: Είναι αναπόφευκτα σφάλματα που οφείλονται στο γεγονός ότι δεν εξετάζεται όλος ο πληθυσμός αλλά μόνο ένα μέρος του (ένα δείγμα). Είναι, δηλαδή, αναπόφευκτη συνέπεια της μεταβλητότητας που υπάρχει στον πληθυσμό. Συνδέονται με την επιλογή σχεδίου δειγματοληψίας, με το μέγεθος του δείγματος και (ασφαλώς) με τη μεταβλητότητα του πληθυσμού.

Η μεταβλητότητα μεταξύ *τυχαίων δειγμάτων*, όπως ήδη αναφέραμε, δεν εξαλείφεται, αλλά η *Θεωρία Πιθανοτήτων* μας επιτρέπει να την περιγράψουμε, ανακοινώνοντας ένα περιθώριο σφάλματος με επιθυμητή πιθανότητα. Έτσι, αν το δείγμα δείξει ότι ένα ποσοστό 45% φοβάται να βγει έξω τη νύχτα και το περιθώριο σφάλματος με πιθανότητα 95% είναι 2% αυτό σημαίνει ότι: με πιθανότητα (εμπιστοσύνη) 95%, το διάστημα μεταξύ 43% και 47% περιέχει το ποσοστό στον πληθυσμό αυτών που φοβούνται να βγουν έξω τη νύχτα⁶.

Για προβληματισμό

- ✓ Στη Νομική επιστήμη η Στατιστική θεωρείται *prima facie*, δηλαδή νομικά ικανή να στηρίζει θεωρία. Ή αλλιώς, για να την αντικρούσεις ως αποδεικτική διαδικασία, πρέπει να χρησιμοποιήσεις αντίστοιχα επιχειρήματα και αντίστοιχες διαδικασίες.
- ✓ Το 1936 έγινε στις Η.Π.Α. (από το περιοδικό *Literary Digest*) μια έρευνα με στόχο την πρόβλεψη του αποτελέσματος των επικείμενων προεδρικών εκλογών. Η έρευνα βασίστηκε στην επιλογή ενός δείγματος 10 εκατομμυρίων πολιτών από τους τηλεφωνικούς καταλόγους των πολιτειών. Στο ερωτηματολόγιο που ταχυδρομήθηκε απάντησε περίπου το 25%. Η ανάλυση των αποτελεσμάτων έδειξε ότι νικητής των εκλογών θα είναι ο υποψήφιος του Ρεπουμπλικανικού κόμματος. Η ημέρα των εκλογών, όμως, επιφύλαξε μια πολύ δυσάρεστη έκπληξη τόσο για τους ερευνητές όσο και για τους οπαδούς του Ρεπουμπλικανικού κόμματος. Ο υποψήφιος του Δημοκρατικού κόμματος (Ρούσβελτ) επικράτησε με ιστορική πλειοψηφία 60%!! Είχε πει ψέματα η Στατιστική ή μήπως κάτι άλλο συνέβη με την έρευνα;
- ✓ Ένα περιοδικό υγιεινής διατροφής θέλει να αποδείξει ότι μεγάλες δόσεις βιταμινών βελτιώνουν την υγεία. Ζήτησε από τους ενήλικες αναγνώστες του που έπαιρναν

⁵ Η έννοια της **πιθανότητας** συνδέεται με φαινόμενα ή πειράματα που είναι μεν μεταβλητά αλλά μακροπρόθεσμα παρουσιάζουν σταθερή συμπεριφορά. Η επιλογή τυχαίου δείγματος είναι τέτοιο φαινόμενο (ανάλογο της ρίψης ενός νομίσματος πολλές φορές).

⁶ Καθόλου ευκαταφρόνητη συνεισφορά από «τα Μαθηματικά της τύχης»! Δε συμφωνείτε;

τακτικά βιταμίνες σε μεγάλες δόσεις, να γράψουν στο περιοδικό περιγράφοντας τις εμπειρίες τους. Από τους 2754 ενήλικες αναγνώστες που απάντησαν το 93% ανέφερε κάποιο όφελος από τη λήψη των βιταμινών. Το ποσοστό αυτό, είναι πιθανόν μεγαλύτερο, μικρότερο ή περίπου το ίδιο με το ποσοστό επί του συνόλου των ενηλίκων που θα παρατηρούσε κάποιο όφελος από μεγάλη λήψη βιταμινών;

- ✓ «Εφόσον ένας νόμος των Μαθηματικών αναφέρεται στην πραγματικότητα δε μπορεί να είναι βέβαιος και αν είναι βέβαιος δε μπορεί να αναφέρεται στην πραγματικότητα» (A. Einstein).
- ✓ «Αβέβαιη γνώση + Γνώση του μεγέθους της αβεβαιότητας σε αυτή = Χρησιμοποίησιμη γνώση» (C.R. Rao).
- ✓ «Τα πιο σπουδαία ερωτήματα στη ζωή είναι στο μεγαλύτερο μέρος τους προβλήματα πιθανοτήτων» (P.-S. Laplace).

Μετρούνται όλα;⁷

Ένα βασικό θέμα κατά τη διαδικασία μελέτης ενός φαινομένου (ή πειράματος) είναι αυτό του προσδιορισμού των μεταβλητών που θα αποτελέσουν το αντικείμενο έρευνας. Πρέπει, δηλαδή, να προσδιοριστεί ποιες από τις πιθανές αιτίες που το προκαλούν θα ελεγχθούν. Φυσικά, με βάση όσα έχουμε ήδη αναφέρει για τη φύση των *τυχαίων φαινομένων/πειραμάτων* (το πολυσύνθετο και το περιορισμένο της γνώσης των αιτίων που τα προκαλούν), γίνεται εύκολα αντιληπτό ότι τέτοια ερωτήματα δεν επιδέχονται απόλυτες απαντήσεις. Αν για παράδειγμα, αποφασίσουμε να μελετήσουμε τη *μεταβλητότητα της ζήτησης ενός προϊόντος*, ως προς την *τιμή του* και ως προς το *ύψος της διαφημιστικής δαπάνης που επενδύεται για την προώθησή του*, αυτό δε σημαίνει ότι μόνο αυτές οι μεταβλητές επηρεάζουν τη μεταβλητότητα της ζήτησης του προϊόντος και ότι άλλες αιτίες όπως η *οικονομική κατάσταση των καταναλωτών*, οι *τιμές ομοειδών προϊόντων* ή η *εποχή του έτους* δεν επηρεάζουν το συγκεκριμένο οικονομικό φαινόμενο. Απλώς, σημαίνει ότι στο πλαίσιο της έρευνας, θα διερευνηθεί η επίδραση μόνο των δύο συγκεκριμένων πιθανών αιτίων διότι αυτά κρίνονται ως τα πιο σημαντικά για τη συγκεκριμένη έρευνα ή ως τα πλέον ενδιαφέροντα ή για κάποιους άλλους λόγους που συνδέονται με τους σκοπούς της έρευνας. Είναι αυτονόητο, ότι σε αυτό το στάδιο, ο ρόλος του ειδικού/ειδικών επί του αντικείμενου της έρευνας, είναι καθοριστικός.

Για κάθε *μεταβλητή* που επιλέγεται για να μελετηθεί, πρέπει να υπάρχει δυνατότητα να της αποδοθούν τιμές. Η απόδοση τιμών σε μια μεταβλητή δε σημαίνει κατ' ανάγκη *ποσοτικοποίηση*, δηλαδή, *μέτρηση ή απαρίθμηση*. Υπάρχουν μεταβλητές όπως το *βάρος*, η *θερμοκρασία* ή ο *αριθμός των μελών μιας οικογένειας* που χαρακτηρίζονται **ποσοτικές**⁸ και οι οποίες, πράγματι, είναι *μετρήσιμες ή απαριθμήσιμες*. Υπάρχουν όμως άλλες, όπως το *άγχος*, η *στάση ως προς κάποιο θέμα*, η *οικογενειακή κατάσταση* ή οι *πολιτικές πεποιθήσεις* που χαρακτηρίζονται **ποιοτικές**⁹ και στις οποίες μπορούν να αποδοθούν τιμές, όμως, οι τιμές αυτές δεν εκφράζουν κάτι το μετρήσιμο ή απαριθμήσιμο αλλά **κατηγοριοποίηση** ή **διάταξη**. Για παράδειγμα, η μεταβλητή *στάση ως προς κάποιο θέμα* μπορεί να πάρει τιμές όπως, *θετική, αδιάφορη, αρνητική*. Σε μια άλλη κλίμακα απόδοσης τιμών θα μπορούσε να πάρει και άλλες τιμές όπως *πολύ θετική, πολύ αρνητική*. Είναι αναπόφευκτο, οι κλίμακες και τα όργανα που χρησιμοποιούνται για την απόδοση τιμών στις *ποιοτικές*

⁷ Ευτυχώς όχι...

⁸ Μια μεταβλητή λέγεται **ποσοτική (quantitative)**, αν παίρνει μόνο αριθμητικές τιμές. Οι ποσοτικές μεταβλητές, ανάλογα με το σύνολο τιμών τους, διακρίνονται σε **συνεχείς (continuous)** (μετρήσιμες) και **διακριτές (discrete)** (απαριθμήσιμες).

⁹ Μια μεταβλητή λέγεται **ποιοτική (qualitative)**, αν δεν παίρνει αριθμητικές τιμές.

μεταβλητές, να έχουν μεγάλο βαθμό υποκειμενισμού. Πώς, εξάλλου, θα ήταν δυνατόν να μετρηθεί αντικειμενικά¹⁰ η ποιότητα¹¹; Τα ερωτηματολόγια και τα τεστ, τα κατ'εξοχήν εργαλεία «μέτρησης» ποιοτικών μεταβλητών, είναι υποκειμενικά εργαλεία «μέτρησης». Αντίθετα, στις ποσοτικές μεταβλητές η μέτρηση ή η απαρίθμηση μπορεί να γίνει αντικειμενικά χωρίς αυτό να σημαίνει ότι στις μετρήσεις ποσοτικών μεταβλητών δεν υπεισέρχονται σφάλματα. Επιδιώξή μας πρέπει να είναι η λήψη μέτρων για τον περιορισμό των σφαλμάτων μέτρησης γιατί αποτελούν πηγή μεροληψίας. Για παράδειγμα, αν τα όργανα μέτρησης είναι απλά, δηλαδή, εύκολα στη χρήση τους, αυτό συμβάλλει στον περιορισμό των σφαλμάτων μέτρησης.

Σε ότι αφορά στις κλίμακες απόδοσης τιμών σε ποιοτικές μεταβλητές, υπάρχουν δύο ευρέως χρησιμοποιούμενες.

1. Κλίμακες κατηγορίας

Στις **κλίμακες κατηγορίας (nominal)** η μόνη σχέση μεταξύ διαφορετικών τιμών είναι η ύπαρξη διαφοράς. Δηλαδή, μια κλίμακα κατηγορίας επιτρέπει μόνο την κατηγοριοποίηση των υποκειμένων (π.χ. διαφορετική κατηγορία χρώματος, διαφορετική κατηγορία πολιτικής προτίμησης).

2. Κλίμακες διάταξης

Στις **κλίμακες διάταξης (ordinal)** μεταξύ διαφορετικών τιμών υπάρχει σχέση διάταξης. Δηλαδή, μια κλίμακα διάταξης, δεν επιτρέπει μόνο την κατηγοριοποίηση των υποκειμένων αλλά και μια σχέση διάταξης-ιεράρχησής τους. Για παράδειγμα, στη μεταβλητή στάση ως προς κάποιο θέμα, μπορούν να αποδοθούν τιμές με βάση την κλίμακα: πολύ θετική, θετική, αδιάφορη, αρνητική, πολύ αρνητική. Επίσης, στη μεταβλητή θέση στην ιεραρχία μιας υπηρεσίας, μπορούν να αποδοθούν τιμές με βάση την κλίμακα: υπάλληλος, προϊστάμενος τμήματος, προϊστάμενος διεύθυνσης, γενικός διευθυντής. Τέλος, στη μεταβλητή αποτελέσματα σε ένα αγώνισμα, μπορούν να αποδοθούν τιμές με βάση την κλίμακα: 1ος, 2ος, 3ος, ... Προφανώς, **ίσες διαφορές μεταξύ τιμών δε συνεπάγονται και ίσες διαφορές στο χαρακτηριστικό** που εκφράζει η μεταβλητή αφού, οι τιμές αυτές δεν ποσοτικοποιούν το χαρακτηριστικό. Η διάταξη εκφράζει π.χ. το «καλύτερο» ή το «προτιμότερο» όχι όμως το «πόσο καλύτερο» ή το «πόσο προτιμότερο». Δηλαδή, 1ος-2ος δεν είναι ίσο με 5ος-6ος.

Σε ότι αφορά τις κλίμακες απόδοσης τιμών σε ποσοτικές μεταβλητές, υπάρχουν δύο επίσης ευρέως χρησιμοποιούμενες.

1. Κλίμακες διαστήματος

Στις **κλίμακες διαστήματος (interval)** γίνεται **ποσοτικοποίηση** του χαρακτηριστικού που εκφράζει η μεταβλητή. Έτσι, μια κλίμακα διαστήματος επιτρέπει, όχι μόνο τη διάταξη-ιεράρχηση των υποκειμένων, αλλά και τον προσδιορισμό επακριβώς της διαφοράς τους. **Ίσες διαφορές μεταξύ τιμών συνεπάγονται και ίσες διαφορές στο χαρακτηριστικό** που εκφράζει η μεταβλητή. Όμως, δεν έχουν νόημα οι αναλογίες. Αυτό συμβαίνει διότι στις κλίμακες διαστήματος μετράμε μεταβλητές στις οποίες δεν εμφανίζεται παντελής έλλειψη του χαρακτηριστικού που εκφράζουν. Δηλαδή, το μηδέν δεν εμφανίζεται εγγενώς στο διάστημα τιμών τους (όπως εμφανίζεται για παράδειγμα στις τιμές της μεταβλητής απόσταση δύο σημείων όπου μηδέν σημαίνει ταύτιση των

¹⁰ Ένα όργανο μέτρησης είναι αντικειμενικό όταν για τη μέτρηση της ίδιας τιμής δίνει το ίδιο αποτέλεσμα οποιοσδήποτε και αν το χρησιμοποιήσει.

¹¹ Η ποιότητα αναγνωρίζεται εύκολα, ορίζεται δύσκολα και είναι αδύνατον να μετρηθεί αντικειμενικά.

σημείων δηλαδή ότι δεν υπάρχει απόσταση), αλλά ορίζεται συμβατικά- αυθαίρετα. Για παράδειγμα, θερμοκρασία 90°C δε συνεπάγεται τριπλάσια αποτελέσματα από θερμοκρασία 30°C γιατί 0°C δε συνεπάγονται παντελή έλλειψη θερμοκρασίας. Έτσι, αν ως μηδέν ορίζαμε τους -273°C , τότε οι 90°C στη νέα κλίμακα αντιστοιχούν σε 363° και οι 30°C σε 303° . Τώρα οι ίδιες θερμοκρασίες δεν έχουν λόγο τρία προς ένα. Όμως, έχει νόημα και στις δυο κλίμακες να πούμε ότι οι δύο θερμοκρασίες διαφέρουν κατά 60° .

Μια ειδική περίπτωση κλίμακας διαστήματος είναι η **κυκλική κλίμακα**. Ένας κύκλος διαιρείται σε 360 μοίρες και οι μηδέν μοίρες ορίζονται αυθαίρετα¹² στο βορρά. Στο βορρά, επίσης, αντιστοιχίζονται οι 360 μοίρες. Σε κυκλική κλίμακα μπορούν να αποδοθούν τιμές π.χ. στις ώρες της ημέρας, στις ενδείξεις της πυξίδας και στους μήνες του έτους¹³. Επίσης, σε κυκλική κλίμακα μετρώνται μεταβλητές που εκφράζουν χαρακτηριστικά **κατεύθυνσης-φοράς**. Τέτοιες μεταβλητές συναντώνται σε πολλές επιστήμες όπως στη Γεωλογία, τη Μετεωρολογία και την Οικολογία (π.χ. η κατεύθυνση του ανέμου, η πορεία των πτηνών, η κατεύθυνση του ίχνους της κίνησης των πάγων, η κατεύθυνση των πτυχώσεων ενός γεωλογικού σχηματισμού π.χ. ψαμμίτη). Τέλος, σε κυκλική κλίμακα μετρώνται μεταβλητές που εκφράζουν χαρακτηριστικά **διεύθυνσης** όπως, η διεύθυνση Landsat γραμμώσεων και η διεύθυνση των αξονικών επιπέδων αντικλίνων, η διεύθυνση γεωλογικών στρωμάτων ή η διεύθυνση των φύλλων των δένδρων. Στις περιπτώσεις αυτές δεν έχει νόημα (ή δεν ενδιαφέρει) η φορά (κατεύθυνση) άλλα μόνο η διεύθυνση. Για παράδειγμα, μπορεί να μας ενδιαφέρει μόνο αν η διεύθυνση-προσανατολισμός ενός φύλλου είναι 10° - 190° και όχι αν η άκρη του φύλλου που στηρίζεται στο μίσχο έχει κατεύθυνση-φορά 10° ή 190° . Βέβαια, τιμές στη διεύθυνση ενός χαρακτηριστικού μπορούν να αποδοθούν με δύο τρόπους. Για παράδειγμα, η διεύθυνση 10° - 190° μπορεί να αποδοθεί είτε ως κατεύθυνση 10° είτε ως κατεύθυνση 190° . Δηλαδή, μια διεύθυνση σε **κυκλική κλίμακα** μπορεί να αποδοθεί ως μια από δύο αντίθετες κατευθύνσεις. Επειδή, το ποια κατεύθυνση επιλέγεται για να αποδοθεί τιμή στην αντίστοιχη διεύθυνση μπορεί να επηρεάσει τα συμπεράσματα της στατιστικής μελέτης (π.χ. μπορεί να διογκωθεί τεχνητά-εσφαλμένα η διασπορά των διευθύνσεων) έχει προταθεί ο εξής μετασχηματισμός των δεδομένων: η τιμή ϕ° της κατεύθυνσης που επιλέγεται (οποιαδήποτε και αν είναι από τις δύο αντίθετες) διπλασιάζεται. Αν η νέα τιμή $2\phi^{\circ}$ βρίσκεται στο διάστημα 0° - 360° τότε αυτή η τιμή χρησιμοποιείται για τη στατιστική επεξεργασία. Αν όχι χρησιμοποιείται η $2\phi^{\circ}$ - 360° . Στη συνέχεια τα αποτελέσματα της επεξεργασίας μετασχηματίζονται αντίστροφα (υποδιπλασιάζονται). Για παράδειγμα, αν για τη διεύθυνση 10° - 190° επιλεγεί να της αποδοθεί η τιμή 10° τότε για τη στατιστική ανάλυση χρησιμοποιείται η τιμή 20° . Αν επιλεγεί η τιμή 190° τότε χρησιμοποιείται η τιμή 380° - $360^{\circ} = 20^{\circ}$. Έτσι, ανεξάρτητα από την κατεύθυνση που κάθε φορά επιλέγεται, για τη διεύθυνση χρησιμοποιείται η ίδια τιμή. Παρατηρήστε ότι

$$\frac{360^{\circ} \phi}{180^{\circ}} = 2\phi.$$

¹² Δεν υπάρχει δηλαδή «φυσική» αιτιολογία.

¹³ x μονάδες του μεγέθους που μετράμε αντιστοιχίζονται (σε κυκλική κλίμακα) σε γωνία $\alpha = (360^{\circ} x)/k$ μοιρών (όπου k ο αριθμός των ίσων διαστημάτων στα οποία διαιρείται ο κύκλος π.χ. 24 για τις ώρες της ημέρας ή 365 για τις ημέρες του έτους). Έτσι, η ώρα 06:15 σε κυκλική κλίμακα αντιστοιχίζεται σε $(360^{\circ} \cdot 6.25)/24 = 93.75$ μοίρες και η ώρα 06:00 σε $(360^{\circ} \cdot 6)/24 = 90$ μοίρες.

Σημείωση 1.2: Επειδή ο κύκλος είναι κλειστή γραμμή (η αρχή – η όποια αρχή- και το τέλος συμπίπτουν), οι συνήθεις μέθοδοι επεξεργασίας δεδομένων, στην πλειονότητά τους, δε μπορούν να εφαρμοσθούν σε δεδομένα κυκλικής κλίμακας γιατί απαιτείται άλλη προσέγγιση (διανυσματική). Για παράδειγμα, η μέση κατεύθυνση των κατευθύνσεων 10^0 και 350^0 , ασφαλώς, δε μπορεί να είναι $(10^0 + 350^0)/2 = 180^0$. Δηλαδή, δύο βόρειες κατευθύνσεις δεν είναι δυνατόν να έχουν μέση κατεύθυνση νότια! Επίσης, η κατεύθυνση 270^0 δε μπορεί να χαρακτηριστεί «μεγαλύτερη» από την κατεύθυνση 90^0 .

2. Κλίμακες αναλογίας

Στις κλίμακες αναλογίας (*ratio*) όπως και στις κλίμακες διαστήματος, γίνεται ποσοτικοποίηση του χαρακτηριστικού που εκφράζει η μεταβλητή. Όμως, οι κλίμακες αναλογίας επιτρέπουν, όχι μόνο τη σύγκριση διαφορών αλλά και τη σύγκριση αναλογιών διότι στο διάστημα τιμών τους το μηδέν περιλαμβάνεται εγγενώς και όχι με αυθαίρετο ορισμό. Έτσι, σε μια κλίμακα αναλογίας, έχει νόημα, δηλαδή απεικονίζει μια πραγματική κατάσταση, τόσο η ισότητα $6 - 3 = 15 - 12$ όσο και η ισότητα $50 = 2 \cdot 25$. Για παράδειγμα, αν το κατάστημα Α έχει τζίρο 15000 € και το κατάστημα Β έχει τζίρο 5000 € τότε έχει νόημα να πούμε και ότι το κατάστημα Α έχει 10000 € τζίρο περισσότερο από το τζίρο του καταστήματος Β και ότι ο τζίρος του καταστήματος Α είναι τριπλάσιος από τον τζίρο του καταστήματος Β.

Ερώτηση: Σε τι κλίμακα μπορούν να αποδοθούν τιμές στις μεταβλητές, αριθμός μητρώου, δαπάνη, άγχος, Δείκτης Τιμών Καταναλωτή, νομός, χρόνος, επίπεδο εκπαίδευσης, βαθμολογία μαθήματος, ηλικία.

Υπάρχουν ψέματα, χοντρά ψέματα και στατιστικές!!!

Η φράση «υπάρχουν τριών ειδών ψέματα: ψέματα, χοντρά ψέματα και στατιστικές¹⁴» που επιλέξαμε ως τίτλο αυτής της ενότητας, συναντάται συχνά στη βιβλιογραφία και αποδίδεται στον Benjamin Disraeli (1804-1881). Οι στατιστικές χρησιμοποιούνται ως συνώνυμο του ψεύδους στον υπερθετικό, μάλιστα, βαθμό. Ασφαλώς ο Βρετανός πρωθυπουργός δεν κυριολεκτεί. Υπαινίσσεται όμως κάτι (δυστυχώς) αδιαμφισβήτητο. Τη διαστρέβλωση, συχνά, της πραγματικότητας από κακή χρήση της Στατιστικής. Δυστυχώς, είτε από άγνοια είτε από σκοπιμότητα, ολοένα και πιο συχνά, γίνεται εσφαλμένη χρήση των στατιστικών μεθόδων παρουσίασης, ανάλυσης και ερμηνείας δεδομένων με αποτέλεσμα να δημιουργείται ψευδής και στρεβλή εικόνα για την πραγματικότητα¹⁵. Τα σχετικά παραδείγματα είναι πολλά. Ας δούμε κάποια από αυτά που συχνά συναντάμε στη βιβλιογραφία ως παραδείγματα κακής χρήσης των στατιστικών μεθόδων.

- Αν οι αριθμοί $x_1 = x_2 = \dots = x_{100} = 1000$ και $x_{101} = 10000000$ εκφράζουν (σε €) το μηνιαίο εισόδημα 101 πολιτών και κάποιος χρησιμοποιήσει τον αριθμητικό μέσο τους

$$\frac{x_1 + x_2 + \dots + x_{100} + x_{101}}{101} = \frac{100 \cdot 1000 + 10000000}{101} = 100000 \text{ €}$$

για να ισχυρισθεί ότι αυτοί οι 101 πολίτες είναι πάρα πολύ καλά αμειβόμενοι, προφανώς προσπαθεί να παραπλανήσει λέγοντας μέρος της «αλήθειας των

¹⁴ There are three kinds of lies: lies, damned lies and statistics.

¹⁵ Είναι γνωστή εξάλλου η φράση “figures do not lie, but liars figure” (οι αριθμοί δεν ψεύδονται αλλά οι ψεύτες αριθμούν).

αριθμών» αφού αποκρύπτει π.χ. ότι η διάμεσος των εισοδημάτων είναι 1000€. Όμως, από τη Στατιστική είναι γνωστό ότι **ο μέσος** ως περιγραφικό μέτρο έχει το μειονέκτημα ότι **είναι ευαίσθητος σε ακραίες τιμές** και γι' αυτό πρέπει να χρησιμοποιείται σε συνδυασμό με άλλα περιγραφικά μέτρα όπως **η διάμεσος**.

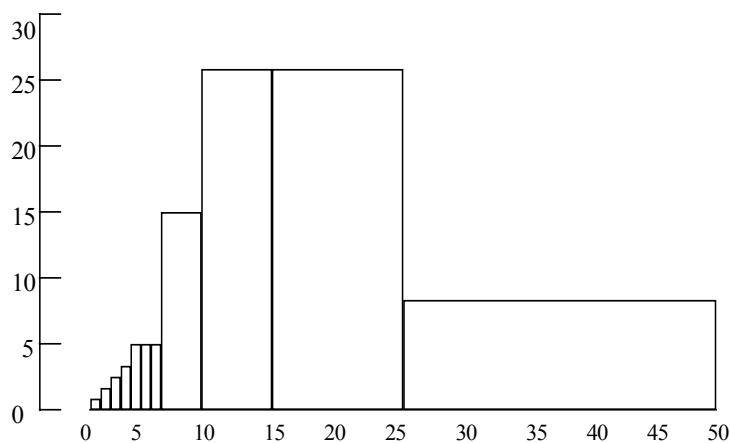
- Στο πλαίσιο των ερευνών για τα αίτια που προκαλούν την εμφάνιση κρουσμάτων πολιομυελίτιδας, είχε παρατηρηθεί ισχυρή θετική συσχέτιση μεταξύ του αριθμού νέων περιστατικών πολιομυελίτιδας και του αριθμού αναψυκτικών που πουλήθηκαν την αντίστοιχη χρονική περίοδο. Είχε δηλαδή παρατηρηθεί ότι όταν υπήρχε μεγάλη κατανάλωση αναψυκτικών, είχαν εκδηλωθεί και πολλά νέα περιστατικά πολιομυελίτιδας. Όταν η κατανάλωση των αναψυκτικών ήταν μειωμένη, υπήρχαν λιγότερα νέα περιστατικά πολιομυελίτιδας. Προκαλούν λοιπόν τα αναψυκτικά εμφάνιση πολιομυελίτιδας; Προφανώς η απάντηση είναι αρνητική. Η επιδημία της πολιομυελίτιδας παρουσιάζει έξαρση το καλοκαίρι, που συμβαίνει να έχουμε και αύξηση της κατανάλωσης των αναψυκτικών. Έτσι εντοπίστηκε ένας τρίτος παράγοντας, *η εποχή του έτους* που είναι καθοριστικός και για τη μεταβλητή *κατανάλωση αναψυκτικών* και για τη μεταβλητή *εμφάνιση κρούσματος πολιομυελίτιδας*. Αν βέβαια οι ερευνητές είχαν συμπεράνει ότι η κατανάλωση αναψυκτικών προκαλεί πολιομυελίτιδα δε θα έφταιγε η Στατιστική. Απλώς οι ερευνητές θα παρέβλεπαν ότι **η συσχέτιση δε συνεπάγεται κατ' ανάγκη αιτιώδη σχέση** και ότι τα αιτιολογικά συμπεράσματα απαιτούν πειραματισμό. Δεν αρκεί η παρατήρηση.
- Στον πίνακα που ακολουθεί φαίνεται η κατανομή των ετήσιων εισοδημάτων των οικογενειών στις Η.Π.Α. το έτος 1973¹⁶.

Ετήσιο εισόδημα (σε χιλιάδες δολάρια)	Ποσοστό οικογενειών (σχετική συχνότητα, %)
0 - 1	1
1 - 2	2
2 - 3	3
3 - 4	4
4 - 5	5
5 - 6	5
6 - 7	5
7 - 10	15
10 - 15	26
15 - 25	26
25 - 50	8

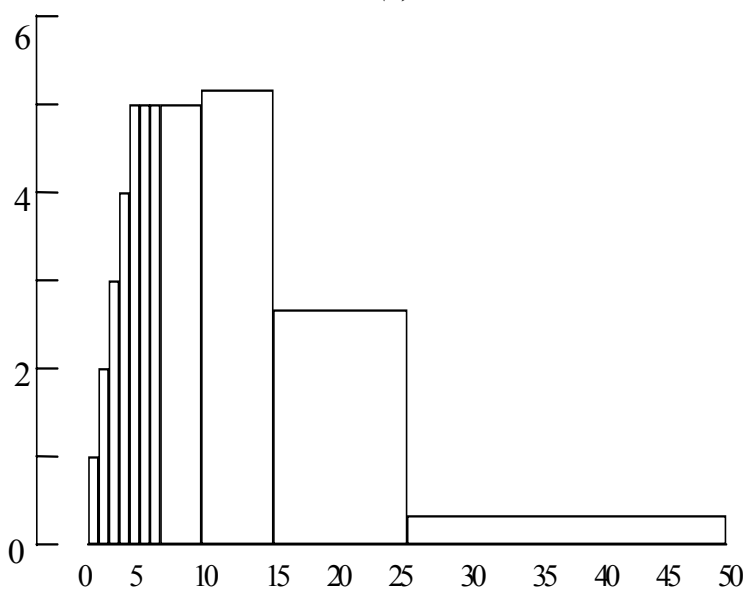
Για τη γραφική αναπαράσταση της κατανομής αυτών των εισοδημάτων, κατασκευάστηκαν τα παρακάτω ιστογράμματα (*Mendenhall et al, 2007*). Παρατηρείστε ότι το ιστογράμμα (**α**) είναι παραπλανητικό διότι δημιουργεί την εντύπωση ότι η οικονομική κατάσταση των οικογενειών στις Η.Π.Α. το έτος 1973 είναι πολύ καλύτερη από την πραγματική. Σύμφωνα με αυτό το ιστογράμμα, οι οικογένειες που έχουν εισόδημα μεγαλύτερο από 25000 δολάρια είναι πολύ περισσότερες από αυτές που έχουν εισόδημα μικρότερο από 7000 δολάρια. Βέβαια, τα πραγματικά δεδομένα (που καταγράφονται στον πίνακα) λένε ακριβώς το αντίθετο (8% και 25% αντίστοιχα). Οι

¹⁶ *Money income in 1973 of families and persons in the United States, Current Population Report, Series P-60 (97), 1975, U.S. Department of Commerce (από το Mendenhall et al, 2007).*

εντυπώσεις όμως μένουν! Φυσικά, δεν ευθύνεται η *Στατιστική*. Ο δημιουργός ενός τέτοιου ιστογράμματος είτε δε γνωρίζει, είτε γνωρίζει αλλά σκοπίμως παραβλέπει το στοιχειώδες ότι στην κατασκευή ενός ιστογράμματος **τα εμβαδά** των ορθογωνίων και **όχι τα ύψη** αντιστοιχούν στις συχνότητες (απόλυτες ή σχετικές). Το σωστό ιστόγραμμα είναι το **(β)** που αποδίδει και τη σωστή εικόνα της κατανομής.



(α)



(β)

Τα τελευταία χρόνια έχει γίνει πολύ εύκολη η πρόσβαση σε προγράμματα υπολογιστών που διευκολύνουν την επεξεργασία δεδομένων και την εφαρμογή στατιστικών μεθόδων. Ο χειρισμός, μάλιστα, αυτών των προγραμμάτων είναι πολύ εύκολος. Το γεγονός αυτό, σε συνδυασμό με το ότι πολλοί νομίζουν ότι η *Στατιστική* αποτελεί εφαρμογή διαδικασιών-συνταγών, έχει ως αποτέλεσμα την όξυνση του προβλήματος της κακής χρήσης της *Στατιστικής*. Είναι μεγάλη πλάνη να πιστεύει κάποιος ότι μπορεί να κάνει *Στατιστική* επειδή έχει πρόσβαση σε ισχυρά υπολογιστικά εργαλεία και δυνατότητα να δημιουργεί με ευκολία π.χ. εντυπωσιακά γραφήματα. Είναι αυτονόητο ότι η σωστή εφαρμογή στατιστικών μεθόδων, πρώτα απ' όλα, απαιτεί την καλύτερη δυνατή γνώση του υπό μελέτη φαινομένου. Απαιτεί επίσης, καλή θεωρητική γνώση ώστε αφενός να επιλεγεί η πλέον κατάλληλη μέθοδος δειγματοληψίας ή να γίνει ο καλύτερος σχεδιασμός του πειράματος και αφετέρου να

επιλεγούν οι καταλληλότερες μέθοδοι παρουσίασης και ανάλυσης των δεδομένων. Κυρίως απαιτεί, σχολαστικό έλεγχο για το αν **ικανοποιούνται οι προϋποθέσεις** εφαρμογής των στατιστικών μεθόδων που επιλέγονται.

Πώς μπορεί να αντιμετωπισθεί το πρόβλημα;

Κατά τη γνώμη μας, το πρόβλημα της κακής χρήσης της *Στατιστικής*, μπορεί να αντιμετωπισθεί, κυρίως, με τη βελτίωση της ποιότητας της παρεχόμενης παιδείας, εκπαίδευσης και κατάρτισης.

Στατιστική και Δεοντολογία

Όπως ήδη έχουμε αναφέρει, είναι πλέον αδιαμφισβήτητο, ότι η *Στατιστική* επηρεάζει πολλές πτυχές της ζωής μας. Είναι επομένως αναμενόμενο, με τη χρήση της *Στατιστικής*, να τίθενται και να εγείρονται διάφορα ηθικά και νομικά ζητήματα. Τα ζητήματα αυτά σχετίζονται τόσο με τη συλλογή των δεδομένων όσο και με την επεξεργασία τους και με τον τρόπο παρουσίασης των αποτελεσμάτων. Ας δούμε κάποια από αυτά.

- ✓ Τα δεδομένα δειγματοληπτικών ή πειραματικών ερευνών επιτρέπεται να δημοσιοποιούνται; Για παράδειγμα, δεδομένα που σχετίζονται με την υγεία, τις πεποιθήσεις, την οικονομική κατάσταση, τις συνήθειες κτλ. των πολιτών επιτρέπεται να δημοσιοποιούνται; Επιτρέπεται, δηλαδή, να είναι διαθέσιμα στον οποιοδήποτε για οποιαδήποτε χρήση και εκμετάλλευση; Στη χώρα μας, όπως συμβαίνει στις περισσότερες πολιτισμένες χώρες, τα δεδομένα δειγματοληπτικών ή πειραματικών ερευνών είναι απόρρητα.
- ✓ Στα *τυχαία συγκριτικά πειράματα* (*randomized comparative experiments*) που γίνονται στο πλαίσιο ερευνών για την αποτελεσματικότητα π.χ. κάποιου νέου φαρμάκου, συγκρίνονται δύο ή περισσότερες *μεταχειρίσεις-επεμβάσεις* (*θεραπευτικές αγωγές*) με την εξέχ, σε γενικές γραμμές, μέθοδο. Επιλέγονται με μια τυχαία διαδικασία δύο ομάδες ασθενών και οι ασθενείς της μιας ομάδας υποβάλλονται σε θεραπευτική αγωγή με το νέο φάρμακο ενώ της άλλης ομάδας όχι (ή υποβάλλονται σε «θεραπεία» με ένα εικονικό φάρμακο). Οι ασθενείς που δεν υποβάλλονται στη θεραπεία αποτελούν την *ομάδα ελέγχου*. Σε τέτοιες πειραματικές έρευνες τίθεται το εξής ζήτημα/δίλημμα. Από τη μια μεριά πρέπει να υπάρχει **επαρκής πίστη** στην αξία του νέου φαρμάκου ώστε να δικαιολογείται η έκθεση των ασθενών της μιας ομάδας στη δράση του και από την άλλη πρέπει να υπάρχει **επαρκής αμφιβολία** για την αποτελεσματικότητά του ώστε να δικαιολογείται η στέρηση της θεραπείας από τους ασθενείς της ομάδας ελέγχου. Ασφαλώς, το δίλημμα είναι πολύ μεγάλο όταν πρόκειται για σοβαρές ασθένειες ή όταν το νέο φάρμακο προκαλεί ισχυρές παρενέργειες. Επίσης, στην ιατρική έρευνα είναι εξίσου σημαντικά τόσο τα θετικά όσο και τα αρνητικά αποτελέσματα. Οι περισσότερες πειραματικές έρευνες αρχίζουν με 50% πιθανότητα επιτυχίας. Αν η πιθανότητα επιτυχίας ήταν πολύ μεγαλύτερη, θα ανησυχούσαμε για τη μη εφαρμογή της θεραπείας σε όλους τους ασθενείς ενώ αν ήταν πολύ μικρότερη θα ανησυχούσαμε για την έκθεση των ασθενών στη δράση της θεραπευτικής αγωγής. Το δίλημμα είναι μεγάλο και **το πρόβλημα παραμένει ανοιχτό**.
- ✓ Πώς πρέπει να ανακοινώνονται τα αποτελέσματα μιας στατιστικής έρευνας; Μπορεί ο τρόπος παρουσιάσής τους να οδηγήσει σε παραπλανητικές ερμηνείες; Αρκεί η παράθεση κάποιων αριθμητικών αποτελεσμάτων; Επιτρέπεται η μερική/ελλιπή/μονομερής/μεροληπτική ανακοίνωση αποτελεσμάτων; Τα

ερωτήματα αυτά και άλλα ανάλογα οδήγησαν τα αναπτυγμένα κράτη στην δημιουργία κωδίκων δεοντολογίας για τον τρόπο ανακοίνωσης των ευρημάτων στατιστικών ερευνών. Έτσι, ταυτόχρονα με την ανακοίνωση των ευρημάτων μιας στατιστικής έρευνας επιβάλλεται, για παράδειγμα, να ανακοινώνονται το δειγματοληπτικό πλαίσιο, η μέθοδος δειγματοληψίας, ο χρόνος διεξαγωγής της έρευνας, το μέγεθος του δείγματος, τα δειγματοληπτικά σφάλματα και άλλες πληροφορίες που συνιστούν την *ταυτότητα της έρευνας*.

- ✓ Πώς εξασφαλίζεται η επαγγελματική επάρκεια, η διαφανής λειτουργία, η αντικειμενικότητα και η ανεξαρτησία από πολιτικές ή άλλες σκοπιμότητες των κρατικών στατιστικών υπηρεσιών (όπως η *Ελληνική Στατιστική Αρχή*); Η σοβαρότητα αυτού του ερωτήματος είναι προφανής αν λάβουμε υπόψη μας ότι οι κρατικές στατιστικές υπηρεσίες συγκεντρώνουν, αναλύουν και ερμηνεύουν δεδομένα με βάση τα οποία λαμβάνονται και τεκμηριώνονται σημαντικές αποφάσεις.

Κλείνουμε αυτή τη σύντομη και κατά το δυνατόν γενική επισκόπηση της *στατιστικής προσέγγισης προβλημάτων*, με ένα απόσπασμα από άρθρο-ομιλία του καθηγητή *Θεόφιλου Κάκουλλου*.

«Οι στατιστικές μέθοδοι απηχούν την προσπάθεια του ανθρώπου να κατανοήσει και να εκφράσει τη γενικότητα των φαινομένων, τα οποία υποπίπτουν στην αντίληψή του και επηρεάζουν κατά το μάλλον ή ήττον τη ζωή του. Μέσα από τον κυκεώνα και το χάος των μεμονωμένων γεγονότων και παρατηρήσεων αναζητούμε συνεχώς κάποια τάξη, όπως αυτή αντανακλάται στη μέση συμπεριφορά, στον ρου των πραγμάτων. Μέσα στην απέραντη ποικιλία και σωρεία των ατομικών χαρακτηριστικών, αναζητούμε τυπικά χαρακτηριστικά, μια καθολική εικόνα του πληθυσμού στον οποίο ανήκουν τα άτομα. Η γενική αυτή εικόνα δεν είναι απλή περίληψη των ατόμων που την απαρτίζουν, αλλά προχωράει πέραν των ατόμων, και αποκτάει νόημα το οποίο δεν θα μπορούσαμε να συλλάβουμε ούτε με την πιο ενδελεχή μελέτη οποιουδήποτε ατόμου».