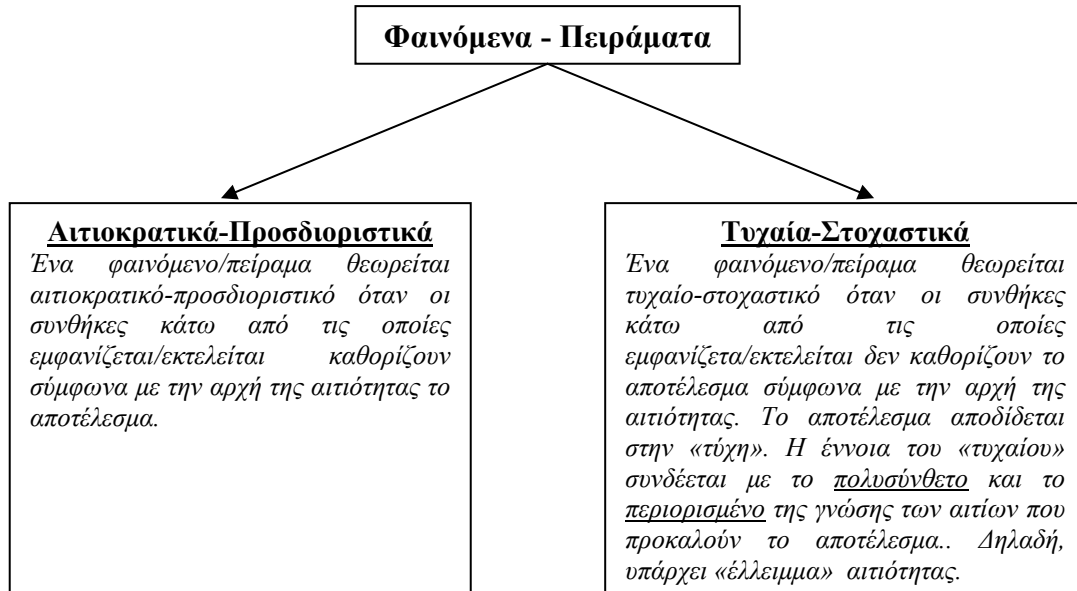


Στατιστική Προσέγγιση Προβλημάτων: Μια Γενική Επισκόπηση

Τα διάφορα προβλήματα (επιστημονικά, κοινωνικά, πολιτικά, κ.λπ.) συνδέονται με φαινόμενα ή με πειράματα¹ τα οποία ταξινομούνται σε δύο γενικές κατηγορίες: Στα **αιτιοκρατικά ή προσδιοριστικά** και στα **τυχαία ή στοχαστικά**.



Για παράδειγμα, αν 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. Μια ασφαλιστική εταιρεία, όμως, 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 επόμενη εβδομάδα σε μια γεωγραφική περιοχή 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/ομοίωση του πραγματικού. Για παράδειγμα, η προσδιοριστική σχέση Κεφάλαιο-Επιτόκιο-Χρόνος-Τόκος μπορεί να περιγραφεί από ένα μαθηματικό τύπο (μοντέλο) 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 και όχι προσδιοριστική αφού η απόδοση της

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 επί των οποίων εξελίσσεται το πείραμα. Αντιθέτως, κατά την παρατήρηση ενός φαινομένου, μετράμε ή παρατηρούμε την κατάσταση των ατόμων ή των αντικειμένων επί των οποίων συμβαίνει το φαινόμενο χωρίς να προσπαθούμε να αλλάξουμε αυτή την κατάσταση με κάποια ειδική μεταχείριση.

καλλιέργειας εξαρτάται εκτός από την ποσότητα λιπάσματος 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 με ακρίβεια πώς επηρεάζουν την 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 επιτρέπουν 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

Σπάνια η πραγματικότητα (φυσικά, κοινωνικά, οικονομικά, πολιτικά, κ.ά φαινόμενα και πειράματα) 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 παραδοχές και «απλούστευση» της πραγματικότητας:

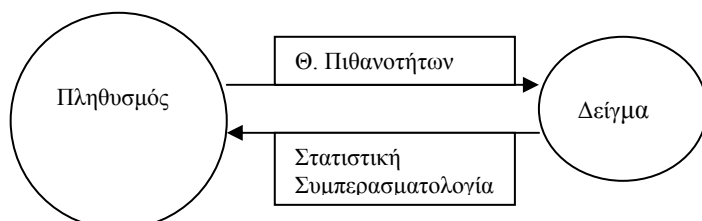
- ✓ Ποιο ποσοστό βλημάτων κατά μέσο όρο πλήττει το στόχο;
- ✓ Πόσα βλήματα πρέπει να χρησιμοποιηθούν για να πληγεί ο στόχος με ικανοποιητική ακρίβεια;
- ✓ Τι μέτρα πρέπει να ληφθούν για να περιορισθεί η διασπορά των σημείων πτώσης των βλημάτων;

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/εξηγήσουμε το φαινόμενο ή το πείραμα και να οδηγηθούμε στην κατασκευή ενός μοντέλου που θα το περιγράφει. 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 μετάβασης από το

² Δεδομένα/Παρατηρήσεις: Στοιχεία που καταγράφονται κατά τη διάρκεια ενός φαινομένου ή κατά την εκτέλεση ενός πειράματος

«όλο» στο «μέρος» (από τον **πληθυσμό**³ στο **δείγμα**) και αντίστροφα, γίνεται, σε γενικές γραμμές, 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**. Η **Στατιστική** μας επιτρέπει 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.



Αξίζει να επισημανθεί ότι κατά τη διαδικασία ελέγχου της συμφωνίας **θεωρίας** και **εμπειρίας**, όπως περιγράφηκε στην προηγούμενη παράγραφο, 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 διαδικασίες λήψης αποφάσεων και συνεπώς επηρεάζει ένα ευρύτατο φάσμα της ανθρώπινης δραστηριότητας⁴.

Σημείωση:

Η στατιστική προσέγγιση που περιγράψαμε, σε γενικές γραμμές παραπάνω, 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

³ **Πληθυσμός:** Είναι σύνολο υποκειμένων (ατόμων, αντικειμένων, τόπων και γενικότερα οποιονδήποτε οντοτήτων) τα οποία ενδιαφερόμαστε να μελετήσουμε ως προς ένα τουλάχιστον κοινό χαρακτηριστικό. Κάθε στοιχείο του πληθυσμού λέγεται **μονάδα (unit)** ή **απλό στοιχείο (element)** και κάθε κοινό χαρακτηριστικό των μονάδων του πληθυσμού λέγεται **μεταβλητή**. **Δείγμα** είναι ένα μέρος του πληθυσμού.

⁴ Ερευνητές, επαγγελματίες, επιχειρηματίες, πολιτικοί, κυβερνήσεις αλλά και κάθε πολίτης στο πλαίσιο των επαγγελματικών τους δραστηριοτήτων αλλά και γενικότερα στην καθημερινή τους ζωή βρίσκονται αντιμέτωποι με 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. Δεδομένα που 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.

Στατιστική Προσέγγιση Προβλημάτων: Μια Γενική Επισκόπηση πρόσφατα, έχουν αναπτυχθεί τεχνικές για τη **Διερευνητική Ανάλυση Δεδομένων (Exploratory Data Analysis)** αφητηρία 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» (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 την αποτελεσματική παρουσίαση των δεδομένων

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, δηλαδή, να μπορούμε να αποφανθούμε με σαφήνεια αν κάποια μονάδα ανήκει σε αυτόν ή όχι. 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: ένας μαθητής ιδιωτικού Γυμνασίου ανήκει στον πληθυσμό; ένας μαθητής της Β' τάξης Γυμνασίου ανήκει στον πληθυσμό; ένας μαθητής νυχτερινού Γυμνασίου ανήκει στον πληθυσμό; Ένας μαθητής ειδικού σχολείου ανήκει στον πληθυσμό; κ.λπ. 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; κ.λπ.). Η κλινική δοκιμή ενός φαρμάκου για τη θεραπεία του έλκους πρέπει, επίσης, να αφορά ένα καλά ορισμένο πληθυσμό (όλα τα άτομα που έχουν έλκος; θα περιορίζεται σε όσα άτομα έχουν ένα συγκεκριμένο τύπο έλκους; μόνο σε όσους χρειάζονται νοσηλεία; κ.λπ.).

Οι πληθυσμοί διακρίνονται σε **πεπερασμένους** και σε **άπειρους**. Με κριτήριο το πλήθος των μονάδων ενός πληθυσμού, φυσικά, άπειροι πληθυσμοί δεν υπάρχουν. Όμως, σε περιπτώσεις πληθυσμών που έχουν μεγάλο πλήθος μονάδων, για τις ανάγκες της πιθανοθεωρητικής επεξεργασίας των δεδομένων, υποθέτουμε ότι αυτά προέρχονται από, πρακτικά, άπειρους πληθυσμούς. Επίσης, θεωρώντας ως πληθυσμό **το σύνολο των τιμών** 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 του πληθυσμού στον οποίο συμβαίνει. Αυτή η μέθοδος συλλογής δεδομένων, η εξέταση δηλαδή όλων των μονάδων του πληθυσμού που μελετάμε, ονομάζεται **ολική απογραφή** ή **απογραφή**. Για παράδειγμα, η Εθνική Στατιστική Υπηρεσία 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 του πληθυσμού, δηλαδή, επιλέγει ένα **δείγμα**, κάνει τις παρατηρήσεις του σε αυτό και στη συνέχεια γενικεύει τα συμπεράσματά του για ολόκληρο τον πληθυσμό. Πώς μπορούμε όμως να επιλέξουμε ένα δείγμα **αντιπροσωπευτικό** του πληθυσμού **ως προς τα χαρακτηριστικά που μελετάμε**; Η απάντηση σε αυτό το ερώτημα δεν είναι εύκολη ούτε μπορεί να είναι απόλυτη. Απαντήσεις σε ερωτήματα σχετικά με τη διαδικασία επιλογής του δείγματος ή με το μέγεθος του δείγματος⁵ δίνει η **Θεωρία Δειγματοληψίας**. Στο πλαίσιο αυτών των σημειώσεων θα αναφερθούμε μόνο στο νόημα της **τυχαίας** και της **απλής τυχαίας δειγματοληψίας**.

Όλοι έχουμε δοκιμάσει μια κουταλιά σούπα για να κρίνουμε ολόκληρο το πιάτο με βάση τη γεύση της. Η σούπα, όμως, είναι ομοιογενής. Η γεύση μιας κουταλιάς αντιπροσωπεύει όλο το υπόλοιπο. 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 αιτιών 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 κ.λπ. Τέτοια όμως δείγματα προκαλούν **μεροληψίες**⁶ δηλαδή, **συστηματικές** διαφορές μεταξύ των αποτελεσμάτων που παίρνουμε από το δείγμα και της πραγματικότητας στον **πληθυσμό** διότι σε τέτοια δείγματα κατά κανόνα

⁵ **Μέγεθος δείγματος** είναι το πλήθος των στοιχείων του.

⁶ **Μεροληψίες ή συστηματικά σφάλματα**: Είναι συστηματικά σφάλματα προς την ίδια κατεύθυνση. Μαζί με τα **μη συστηματικά ή τυχαία σφάλματα** αποτελούν τα **μη δειγματοληπτικά σφάλματα**. Πηγές μη δειγματοληπτικών σφαλμάτων είναι, μεταξύ άλλων, η λάθος επιλογή δείγματος, τα λάθη στα ερωτηματολόγια, τα λάθη συνέντευξης, τα λάθη επεξεργασίας κ.ά. Τέτοιου είδους σφάλματα εμφανίζονται και στις απογραφές. Από αυτά, τα μη συστηματικά (τυχαία) αλληλοαναιρούνται (ιδιαίτερα στα μεγάλα δείγματα). Για τις μεροληψίες, δεν υπάρχει κάποιος γενικός κανόνας εκτίμησης του μεγέθους τους.

περιλαμβάνονται άτομα συγκεκριμένων κατηγοριών και αποκλείονται άτομα όλων των άλλων. Για παράδειγμα, μπορεί να αισθανόμαστε πιο ασφαλείς να πλησιάσουμε καλοντυμένα, μεσαίου εισοδήματος άτομα αλλά έτσι οδηγούμαστε στο να αποκλείσουμε από το δείγμα 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.

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*.

Τυχαία δειγματοληψία είναι η μέθοδος δειγματοληψίας σύμφωνα με την οποία στην επιλογή του δείγματος καταλήγουμε με δεδομένη πιθανότητα. Στην περίπτωση αυτή το δείγμα λέγεται *τυχαίο*. 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 εκλεγεί. Δηλαδή, αν από ένα πληθυσμό μπορούμε να πάρουμε *k* το πλήθος δείγματα μεγέθους *n*, τότε $p_1 = p_2 = p_3 = \dots = p_k = \frac{1}{k}$ (όπου, $p_i, i = 1, 2, \dots, k$ η πιθανότητα να εκλεγεί το δείγμα *i*). Ένα δείγμα που επιλέγεται με τέτοιο τρόπο λέγεται *απλό τυχαίο δείγμα*.

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 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 βασιστούμε 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 (stratified sampling)*, η *συστηματική (systematic sampling)*, η *κατά συστάδες (cluster sampling)* και η *δισταδιακή κατά συστάδες (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 πιθανοθεωρητική⁷ εκτίμηση του μεγέθους των δειγματοληπτικών σφαλμάτων⁸ αλλά και τον περιορισμό τους. 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 νύχτα⁹.

⁷ 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).

⁸ *Δειγματοληπτικά σφάλματα*: Αναπόφευκτα σφάλματα που συνδέονται με την επιλογή σχεδίου δειγματοληψίας και με το μέγεθος και τη μεταβλητότητα του δείγματος.

⁹ Καθόλου ευκαταφρόνητη συνεισφορά από «τα Μαθηματικά της τύχης»! Δε συμφωνείτε; Εργαστήριο Μαθηματικών&Στατιστικής/Γ. Παπαδόπουλος (www.aua.gr/gpapadopoulos)

Για προβληματισμό:

- ✓ Στη Νομική Επιστήμη η Στατιστική θεωρείται *prima facie*, δηλαδή νομικά ικανή να στηρίξει 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 10.000.000 πολιτών από τους τηλεφωνικούς καταλόγους των διαφόρων 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 τακτικά βιταμίνες σε μεγάλες δόσεις, να γράψουν στο περιοδικό περιγράφοντας τις εμπειρίες τους. Από τους 2.754 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 όλων των ενηλίκων που θα παρατηρούσαν κάποιο όφελος από μεγάλη λήψη βιταμινών;
- ✓ Πώς σχολιάζετε τις φράσεις:

«Εφόσον ένας νόμος των Μαθηματικών αναφέρεται στην πραγματικότητα δε μπορεί να είναι βέβαιος και αν είναι βέβαιος δε μπορεί να αναφέρεται στην πραγματικότητα».

«Αβέβαιη γνώση + Γνώση του μεγέθους της αβεβαιότητας σε αυτή = Χρησιμοποιήσιμη γνώση».

«Τα πιο σπουδαία ερωτήματα στη ζωή είναι, στο μεγαλύτερο μέρος τους, προβλήματα πιθανοτήτων».

Μετριούνται όλα;¹⁰

Αναφερθήκαμε, ήδη, στην έννοια της μεταβλητής. Σε αυτήν την παράγραφο θα αναφερθούμε στη «μέτρηση» των μεταβλητών.

Ένα βασικό θέμα κατά τη διαδικασία της 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

¹⁰ *Ευτυχώς όχι!*

να μελετήσουμε τη μεταβλητότητα της ζήτησης ενός προϊόντος, ως προς την τιμή του προϊόντος και ως προς το επίπεδο της σχετικής διαφημιστικής δαπάνης, 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 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 αντικειμένου 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**¹¹ 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**¹² 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 μεταβλητές**, να έχουν μεγάλο βαθμό υποκειμενισμού. Πώς, εξάλλου, θα ήταν δυνατόν να μετρηθεί **αντικειμενικά**¹³ η **ποιότητα**¹⁴; 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. Επιδίωξή 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 (nominal)**

Στις **κλίμακες κατηγορίας**, 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 (ordinal)**

Στις **κλίμακες διάταξης**, 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

¹¹ Μια μεταβλητή λέγεται **ποσοτική**, αν παίρνει μόνο αριθμητικές τιμές. Οι ποσοτικές μεταβλητές, ανάλογα με το σύνολο τιμών τους, διακρίνονται σε **συνεχείς** (μετρήσιμες) και **διακριτές** (απαριθμήσιμες).

¹² Μια μεταβλητή λέγεται **ποιοτική**, αν δεν παίρνει αριθμητικές τιμές.

¹³ Ένα όργανο μέτρησης είναι **αντικειμενικό** όταν δίνει το ίδιο αποτέλεσμα για τη μέτρηση της ίδιας τιμής οποιοσδήποτε και αν το χρησιμοποιήσει.

¹⁴ Η **ποιότητα** αναγνωρίζεται εύκολα, ορίζεται δύσκολα, είναι, όμως, αδύνατον να μετρηθεί αντικειμενικά..

Στατιστική Προσέγγιση Προβλημάτων: Μια Γενική Επισκόπηση
 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.

Επίσης, υπάρχουν δύο, ευρέως χρησιμοποιούμενες, κλίμακες απόδοσης τιμών σε ποσοτικές μεταβλητές:

1. Κλίμακες διαστήματος (interval)

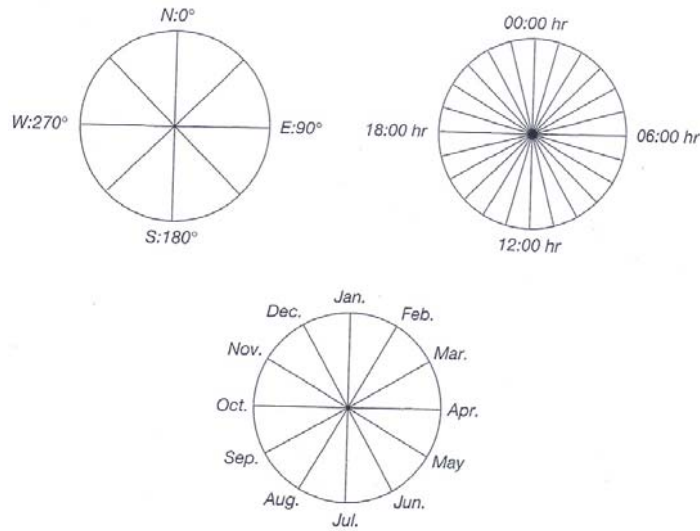
Στις **κλίμακες διαστήματος**, γίνεται **ποσοτικοποίηση** του χαρακτηριστικού που εκφράζει η μεταβλητή. Έτσι, μια κλίμακα διαστήματος επιτρέπει, όχι μόνο τη διάταξη-ιεράρχηση των υποκειμένων, αλλά και τον προσδιορισμό επακριβώς της διαφοράς τους. **Ίσες διαφορές μεταξύ τιμών συνεπάγονται και ίσες διαφορές 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 σημείων δηλαδή ότι δεν υπάρχει απόσταση), αλλά ορίζεται συμβατικά-αυθαίρετα. Για παράδειγμα, θερμοκρασία 90⁰C δε συνεπάγεται τριπλάσια αποτελέσματα από θερμοκρασία 30⁰C γιατί 0⁰C δε συνεπάγονται παντελή έλλειψη θερμοκρασίας. Έτσι, αν ως μηδέν ορίζαμε τους -273⁰C, τότε οι 90⁰C στη νέα κλίμακα αντιστοιχούν σε 363⁰ και οι 30⁰C σε 303⁰. 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 60⁰. Κλίμακα διαστήματος είναι επίσης η κλίμακα επιπέδου υπηρεσιών (γιατί;).

Μια ειδική περίπτωση κλίμακας διαστήματος είναι η **κυκλική κλίμακα**. Ένας κύκλος διαιρείται σε 360 μοίρες και οι μηδέν μοίρες ορίζονται αυθαίρετα¹⁵ 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 έτους¹⁶.

¹⁵ Δεν υπάρχει δηλαδή «φυσική» αιτιολογία

¹⁶ x μονάδες του μεγέθους που μετράμε αντιστοιχίζονται (σε κυκλική κλίμακα) σε γωνία $\alpha = \frac{360^\circ \cdot x}{k}$ 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 $\frac{360^\circ \cdot 6.25}{24} = 93.75$ μοίρες και η

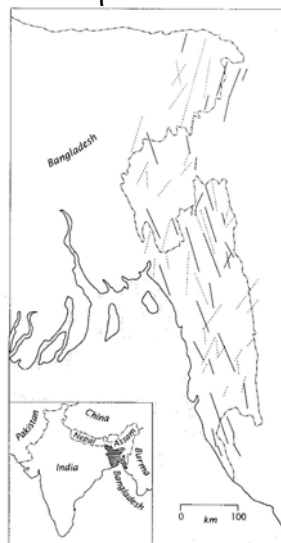
ώρα 06:00 σε $\frac{360^\circ \cdot 6}{24} = 90$ μοίρες.



Επίσης, σε κυκλική κλίμακα μετρώνται μεταβλητές που εκφράζουν χαρακτηριστικά **κατεύθυνσης-φοράς (directional)**. 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 (βλ. ακόλουθο σχήμα), η κατεύθυνση των πτυχώσεων ενός γεωλογικού σχηματισμού π.χ. ψαμμίτη).



Σε κυκλική κλίμακα μετρώνται, επίσης, μεταβλητές που εκφράζουν χαρακτηριστικά **διεύθυνσης (oriented, axial)** όπως η διεύθυνση Landsat γραμμώσεων και η διεύθυνση των αξονικών επιπέδων αντικλίνων (βλ. ακόλουθο σχήμα), 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 10^0 - 190^0 και όχι αν η άκρη του φύλλου που στηρίζεται στο μίσχο έχει κατεύθυνση-φορά 10^0 ή 190^0 .

Είναι προφανές ότι σε *κυκλική κλίμακα*, μπορούν να αποδοθούν τιμές στη διεύθυνση ενός χαρακτηριστικού με δύο τρόπους. Για παράδειγμα, η διεύθυνση 10^0 - 190^0 μπορεί να αποδοθεί είτε ως κατεύθυνση 10^0 είτε ως κατεύθυνση 190^0 . 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^0 της κατεύθυνσης που επιλέγεται (οποιαδήποτε και αν είναι από τις δύο αντίθετες) διπλασιάζεται. Αν η νέα τιμή $2\phi^0$ βρίσκεται στο διάστημα 0^0 - 360^0 τότε αυτή η τιμή χρησιμοποιείται για τη στατιστική επεξεργασία. Αν όχι χρησιμοποιείται η $2\phi^0$ - 360^0 . Στη συνέχεια τα αποτελέσματα της επεξεργασίας μετασχηματίζονται αντίστροφα (υποδιπλασιάζονται). Για παράδειγμα, αν για τη διεύθυνση 10^0 - 190^0 επιλεγεί να της αποδοθεί η τιμή 10^0 τότε για τη στατιστική ανάλυση χρησιμοποιείται η τιμή 20^0 . Αν επιλεγεί η τιμή 190^0 τότε χρησιμοποιείται η τιμή 380^0 - $360^0 = 20^0$. Έτσι, ανεξάρτητα από την κατεύθυνση που κάθε φορά επιλέγεται, για τη διεύθυνση χρησιμοποιείται η ίδια τιμή.

Παρατηρείστε ότι: $\frac{360^0 \cdot \phi}{180^0} = 2 \cdot \phi$.

Σημείωση:

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 , φυσικά, δε μπορεί να είναι $\frac{10^0 + 350^0}{2} = 180^0$. 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 (ratio)

Στις *κλίμακες αναλογίας*, 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 15.000 € και το κατάστημα Β έχει τζίρο 5.000 € τότε έχει νόημα να πούμε και ότι το κατάστημα Α έχει 10.000 € τζίρο περισσότερο από το τζίρο του καταστήματος Β και ότι ο τζίρος του καταστήματος Α είναι τριπλάσιος από τον τζίρο του καταστήματος Β.

Ερωτήσεις:

- Η κλίμακα θερμοκρασίας Kelvin είναι κλίμακα διαστήματος ή κλίμακα αναλογίας;

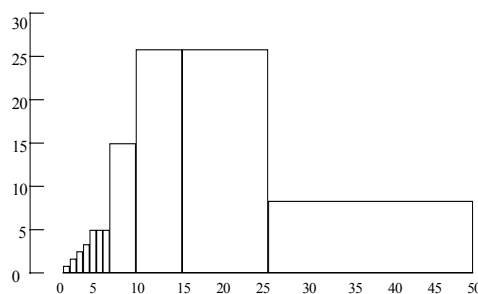
- Σε τι κλίμακα μπορούν να αποδοθούν τιμές στις μεταβλητές: Αριθμός μητρώου, Δαπάνη, Άγχος, Δείκτης Τιμών Καταναλωτή, Νομός, Χρόνος, Επίπεδο εκπαίδευσης, Βαθμολογία μαθήματος, Ηλικία

Υπάρχουν ψέματα, μεγάλα ψέματα και η Στατιστική!!!

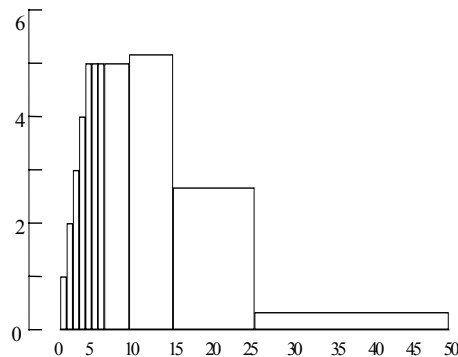
Ο τίτλος αυτής της παραγράφου έχει χρησιμοποιηθεί ως τίτλος βιβλίου¹⁷. Ο συγγραφέας χρησιμοποιεί τη λέξη *Στατιστική* ως τον υπερθετικό βαθμό της λέξης *ψέμα*! Φυσικά, δεν κυριολεκτεί. Υπαινίσσεται, όμως, ένα αδιαμφισβήτητο γεγονός. Τη διαστρέβλωση, πολλές φορές, της πραγματικότητας που γίνεται από κακή χρήση της *Στατιστικής*. Δυστυχώς, από άγνοια ή 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 πραγματικότητα¹⁸. Τα σχετικά παραδείγματα είναι πολλά. Ας δούμε τρία τέτοια παραδείγματα:

(I) Τα παρακάτω ιστογράμματα κατασκευάστηκαν για να περιγράψουν την κατανομή των οικογενειών στις Η.Π.Α. ως προς το ετήσιο εισόδημά τους το έτος 1973. Το ιστόγραμμα (α) είναι παραπλανητικό διότι δημιουργεί την εντύπωση ότι η οικονομική κατάσταση των οικογενειών στις Η.Π.Α. είναι πολύ καλύτερη από την πραγματική. Σύμφωνα με αυτό το ιστόγραμμα, οι οικογένειες που έχουν εισόδημα μεγαλύτερο από 25.000 δολάρια είναι πολύ περισσότερες από αυτές που έχουν εισόδημα μικρότερο από 7.000 δολάρια. Βέβαια, τα πραγματικά δεδομένα λένε ακριβώς το αντίθετο (8% και 25% αντίστοιχα). Οι εντυπώσεις όμως μένουν! Φυσικά, δεν ευθύνεται η *Στατιστική*. Ο δημιουργός αυτού του ιστογράμματος είτε δε γνώριζε, είτε γνώριζε, αλλά σκοπίμως αγνόησε το στοιχειώδες ότι στην κατασκευή ενός ιστογράμματος **τα εμβαδά** των ορθογωνίων και **όχι τα ύψη** αντιστοιχούν στις συχνότητες (απόλυτες ή σχετικές αναλόγως). Το σωστό ιστόγραμμα είναι το (β) που αποδίδει και τη σωστή εικόνα της κατανομής.

(α)



(β)



¹⁷ *There are lies, big lies and statistics*

¹⁸ Είναι γνωστή εξάλλου η φράση “figures do not lie, but liars figure” (οι αριθμοί δεν ψεύδονται αλλά οι ψεύτες αριθμούν)

| Ετήσιο εισόδημα σε χιλιάδες δολάρια | Ποσοστό οικογενειών (σχετική συχνότητα f_i %) |
|-------------------------------------|---|
| 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 |

(II) Στην Αμερική πριν εισαχθεί το εμβόλιο Salk κατά της πολιομυελίτιδας, ερευνητές εξέταζαν αν υπάρχει σχέση ανάμεσα στην εμφάνιση της πολιομυελίτιδας και στον αριθμό των πωληθέντων αναψυκτικών. Για κάθε εβδομάδα του έτους κατέγραφαν σε ένα πίνακα τον αριθμό των αναψυκτικών που καταναλώθηκαν τη συγκεκριμένη εβδομάδα και τον αριθμό των νέων περιστατικών πολιομυελίτιδας που είχαν αναφερθεί. Τα δεδομένα αυτά εμφάνιζαν ισχυρή θετική συσχέτιση ανάμεσα στον αριθμό των περιστατικών της πολιομυελίτιδας και στον αριθμό των πωληθέντων αναψυκτικών. Τις εβδομάδες που είχαν καταναλωθεί περισσότερα αναψυκτικά, είχαν εκδηλωθεί περισσότερα νέα περιστατικά πολιομυελίτιδας. Όταν η κατανάλωση των αναψυκτικών ήταν μειωμένη, υπήρχαν λιγότερα νέα περιστατικά. Προκαλούν λοιπόν τα αναψυκτικά εμφάνιση πολιομυελίτιδας; Προφανώς η απάντηση είναι αρνητική. Αν ήταν έτσι με την απαγόρευση της πώλησης τους θα έπρεπε να μειώνεται και η εμφάνιση της νόσου. 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 μεταβλήτη κατανάλωση αναψυκτικών και για τη μεταβλήτη κρούσμα πολιομυελίτιδας. Αν βέβαια οι ερευνητές είχαν συμπεράνει ότι η κατανάλωση αναψυκτικών προκαλεί πολιομυελίτιδα δεν θα έφταιγε η Στατιστική. Απλώς οι ερευνητές αγνοούσαν βασικά θεωρητικά θέματα της Στατιστικής όπως ότι η **συσχέτιση δε συνεπάγεται κατ' ανάγκη αιτιώδη σχέση** και ότι τα αιτιολογικά συμπεράσματα απαιτούν πειραματισμό. Δεν αρκεί η παρατήρηση.

(III) Αν οι αριθμοί $x_1 = x_2 = \dots = x_{100} = 1.000$ και $x_{101} = 10.000.000$ δείχνουν το εισόδημα 101 φορολογουμένων σε € και κάποιος ισχυρισθεί ότι το μέσο εισόδημα τους ανέρχεται σε 100.000 € προφανώς προσπαθεί να παραπλανήσει λέγοντας μέρος της «αλήθειας των αριθμών». Σκόπιμα ή από άγνοια αποκρύπτει τη διάμεσο των εισοδημάτων (που είναι 1.000 €). Όμως, από τη Στατιστική είναι γνωστό ότι η **μέση τιμή** ως περιγραφικό μέτρο έχει το μειονέκτημα ότι **είναι πολύ ευαίσθητη σε ακραίες τιμές** και γι' αυτό πρέπει να χρησιμοποιείται σε συνδυασμό με άλλα περιγραφικά μέτρα όπως η **διάμεσος**.

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 είναι άκριτη 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.

Δεοντολογία και Στατιστική

Όπως ήδη έχουμε αναφέρει, είναι πλέον αδιαμφισβήτητο, ότι η *Στατιστική* επηρεάζει πολλές πτυχές της ζωής μας. Είναι επομένως φυσικό, με τη χρήση της *Στατιστικής*, να εγείρονται πολλά 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 αναφερθούμε σε 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 κ.λπ. των πολιτών επιτρέπεται να δημοσιοποιούνται; Επιτρέπεται, δηλαδή, να είναι διαθέσιμα στον οποιονδήποτε για οποιαδήποτε χρήση και εκμετάλλευση; Στη χώρα μας, όπως συμβαίνει στις περισσότερες πολιτισμένες χώρες, τα δεδομένα δειγματοληπτικών ή πειραματικών ερευνών είναι απόρρητα (π.χ. αρθρ. 40 του Ν.Δ. 3627/56).
- ✓ 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 εξέξ, σε γενικές γραμμές, μέθοδο: Επιλέγονται, για παράδειγμα, δύο τυχαία δείγματα ασθενών του ίδιου μεγέθους. Οι ασθενείς του ενός υποβάλλονται σε θεραπευτική αγωγή με το νέο φάρμακο και του άλλου όχι (ή υποβάλλονται σε «θεραπεία» με ένα εικονικό φάρμακο). Οι ασθενείς που δεν υποβάλλονται στη θεραπεία αποτελούν την *ομάδα ελέγχου*.

Υπάρχει μια λεπτή ισορροπία μεταξύ του αν πρέπει να κάνουμε ή να μην κάνουμε μια τέτοια πειραματική έρευνα. Από τη μια μεριά, πρέπει να υπάρχει **επαρκής πίστη** στην αξία του νέου φαρμάκου ώστε να δικαιολογείται η έκθεση των ασθενών του ενός δείγματος στη δράση αυτού του φαρμάκου. Από την άλλη μεριά, πρέπει να υπάρχει **επαρκής αμφιβολία** για τη δραστηριότητά του ώστε να δικαιολογείται η στέρηση της θεραπείας από τους ασθενείς της ομάδας ελέγχου (φυσικά, το ηθικό δίλημμα είναι πολύ μεγάλο όταν πρόκειται για σοβαρές ασθένειες ή όταν το νέο φάρμακο προκαλεί ισχυρές παρενέργειες). Επιπλέον, για την απόκτηση βάσιμης ιατρικής μαρτυρίας, 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/ελλειπής/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 Εθνική Στατιστική Υπηρεσία Ελλάδος); 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 οικονομικές και πολιτικές αποφάσεις. Για παράδειγμα, η τιμή του *Δείκτη Τιμών Καταναλωτή* επηρεάζει μια σειρά κυβερνητικών αποφάσεων, όπως, τις αποφάσεις για τα κλιμάκια φορολογίας, την εισοδηματική πολιτική κ.ά. Επίσης, η τιμή του *ποσοστού ανεργίας*, εκτός των άλλων, μπορεί να προκαλέσει ακόμη και πολιτικές ανακατατάξεις.

Κλείνουμε αυτή τη σύντομη, και κατά το δυνατόν, γενική επισκόπηση της *στατιστικής προσέγγισης των φαινομένων*, 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».