



Κεφάλαιο

1

# Η Επισκόπηση της Στατιστικής

(Overview of Statistics)

## 1.1 Ορισμός της Στατιστικής Επιστήμης

Η Στατιστική είναι ο κλάδος της επιστήμης ο οποίος ασχολείται με τις μεθόδους που είναι κατάλληλες για την συλλογή, την οργάνωση, την παρουσίαση και την ανάλυση δεδομένων, όπως επίσης, για τη χρησιμοποίηση των πληροφοριών που περιέχουν τα δεδομένα, με το σκοπό να συνταχθούν *συμπεράσματα* και να ληφθούν *αποφάσεις*, ειδικά όταν υπάρχει το στοιχείο της *αβεβαιότητας*.

Η Στατιστική ορίζεται επίσης σαν «η επιστήμη της λήψης αποφάσεων όταν υπάρχει η αβεβαιότητα».

Οι μέθοδοι της Στατιστικής χρησιμοποιούνται σε πολλά πεδία, όπως στην Τεχνική, στο Εμπόριο, στη Βιολογία, στην Ιατρική, Γεωργία, Ψυχολογία και Εκπαίδευση. Πρέπει να τονισθεί ότι οι μέθοδοι της Στατιστικής βοηθούν μόνο στη λήψη αποφάσεων. Πολύ σπάνια οι μέθοδοι αυτές πρέπει να χρησιμοποιούνται σαν μοναδική βάση για μία απόφαση.

Η Στατιστική παρουσιάζει στον λαμβάνοντα τις αποφάσεις, είτε αυτός είναι διευθυντής σε μία επιχείρηση, είτε είναι τεχνικός, είτε επιστήμονας, ορισμένα γεγονότα, και σε πολλές περιπτώσεις, δίνει έναν εκτιμητή της πιθανότητας να ληφθεί η λανθασμένη απόφαση, όπως επίσης, τις πιθανές χρηματικές συνέπειες της λήψης της λανθασμένης απόφασης. Στην Τεχνική, στον Εμπορικό κόσμο και στην Επιστήμη γενικά, οι ιδέες, τεχνικές και τα αποτελέσματα της Στατιστικής είναι απαραίτητα συστατικά στοιχεία στη διαδικασία λήψης αποφάσεων.

## 1.2 Ιστορία της Στατιστικής

Η λέξη «Στατιστική» παράγεται από την λατινική λέξη «Status» που σημαίνει «κράτος».

Το φαινόμενο της ανάπτυξης των Μαθηματικών της θεωρίας Πιθανοτήτων, συνέβαλε κατά πολύ στην ανάπτυξη της σύγχρονης Στατιστικής. Επίσης, η ανάγκη των κυβερνήσεων να συλλέγουν πληροφορίες για τις βασικές δραστηριότητες των πολιτών στη Γεωργία και το Εμπόριο, όπως επίσης και για τον αριθμό των γεννήσεων και των θανάτων, ήταν άλιος ένας παράγοντας ανάπτυξης της Στατιστικής.

Η συλλογή δεδομένων γινόταν σε όλη τη διάρκεια της γραπτής ιστορίας. Κατά την διάρκεια του Αιγυπτιακού, Ελληνικού και Ρωμαϊκού πολιτισμού, τα δεδομένα που αφορούσαν τους πολίτες μαζεύονταν για τον σκοπό της επιβολής φορολογίας και στρατιωτικής θητείας.

Στους Βυζαντινούς χρόνους, συνήθως, τα μοναστήρια και οι διευθύνσεις εκκλησιαστικών περιφερειών κρατούσαν στοιχεία που αφορούσαν γεννήσεις, γάμους, θανάτους κ.τλ.

Στην Αμερική, διάφορα δεδομένα συλλέχθηκαν κατά τους χρόνους της αποικιοκρατίας και το 1970 το Ομοσπονδιακό Σύνταγμα καθιέρωσε την διεξαγωγή απογραφής κάθε 10 χρόνια.

Φημισμένοι πιθανοθεωρητικοί και στατιστικοί επιστήμονες είναι, οι Άγγλοι, R. Fisher, K. Pearson, M. Stone, ο Πολωνός J. Newman, οι Ρώσοι Kolmogorov, Chebyshev, Markov, οι Αμερικανοί Tukey, B. Efron, ο Γερμανός R. Von Mises και άλλοι.

### **1.3 Η Επίδραση των Ηλεκτρονικών Υπολογιστών στην Εξαγωγή Στατιστικών Συμπερασμάτων**

Ο ηλεκτρονικός υπολογιστής έχει αυξήσει κατά πολύ τη χρήση των στατιστικών μεθόδων. Επειδή ο Η/Υ μπορεί να κάνει εκατομμύρια υπολογισμούς σε δευτερόλεπτα, κατέστησε εφικτή τη χρήση στατιστικών μεθόδων σε πεδία που αυτό ήταν προηγουμένως αδύνατο.

Μερικά από τα σπουδαιότερα στατιστικά πακέτα που χρησιμοποιούνται σήμερα σε μικροϋπολογιστές και μεγάλες μονάδες υπολογιστών, για την στατιστική ανάλυση δεδομένων, είναι τα εξής:

- STATGRAPHICS
- Statistical Package for the Social Sciences (SPSS)
- MICROSTAT
- MINITAB
- GB - STAT
- S-Plus
- SAS
- Numerical Algorithms Group for Statistics (NAG)
- SSP

## 1.4 Ο Ρόλος της Στατιστικής στην Ανάπτυξη των άλλων Επιστημών

Κατ' αρχή θα δώσουμε τους ακόλουθους δύο ορισμούς για να εκθέσουμε σαφέστερα τον ρόλο της Στατιστικής στην ανάπτυξη της σύγχρονης επιστήμης:

### Ορισμός 1.4.1

*Στατιστικός Πληθυσμός* είναι το σύνολο όλων των προσώπων, αντικειμένων και παρατηρήσεων με ένα ιδιαίτερο χαρακτηριστικό.

Παραδείγματα στατιστικών πληθυσμών είναι: α) Όλα τα ήλυστα των αυτοκινήτων που κατασκευάστηκαν από μία εταιρεία κατά τη διάρκεια του προηγούμενου έτους. β) Το σύνολο όλων των προσώπων που εμβολιάστηκαν με ένα ειδικό εμβόλιο το 2015.

### Ορισμός 1.4.2

*Δείγμα* λέγεται ένα υποσύνολο ενός στατιστικού πληθυσμού. Τα στοιχεία του δείγματος εκλέγονται από τον πληθυσμό.

Ένα παράδειγμα δείγματος είναι το εξής: Τα ονόματα 10 σπουδαστών που εκλέγονται τυχαία από μία τάξη 100 σπουδαστών οι οποίοι διδάσκονται το μάθημα της Στατιστικής.

Η πειραματική επιστημονική έρευνα στη Μηχανική και στην επιστήμη γενικότερα, περιλαμβάνει τη χρήση πειραματικών δεδομένων που περιέχονται σε ένα δείγμα, από το οποίο, μετά από στατιστική ανάλυση, συμπεραίνεται η φύση του πληθυσμού που χαρακτηρίζει το φαινόμενο το οποίο είναι αντικείμενο έρευνας του ερευνητή.

As εξετασθεί ο τρόπος με τον οποίο η επιστήμη της Στατιστικής συνεισφέρει στην ανωτέρω διαδικασία: Κατ' αρχήν πρέπει να τονισθεί ότι, τα συμπεράσματα που βασίζονται σε δειγματικά δεδομένα σχεδόν πάντα θα

υπόκεινται σε σφάλμα, γιατί ένα δείγμα δεν είναι η ακριβής εικόνα ενός πληθυσμού. Η φύση των πληροφοριών που δίνονται, εξαρτάται από το ειδικότερο δείγμα που εκλέχθηκε, και συνεπώς θα αλλιάζει από δείγμα σε δείγμα.

Ας υποθεθεί ότι ζητείται να εκτιμηθεί το ποσοστό των ελλειψοειδών ηλεκτρικών λαμπτήρων που κατασκευάζονται σε ένα εργοστάσιο. Υποθέτουμε ότι εξετάζεται η ποιότητα 100 λαμπτήρων και ευρίσκονται 47 ελλειψοειδή. Μήπως αυτό σημαίνει ότι, ακριβώς το 47% όλων των λαμπτήρων που κατασκευάζονται από την παραπάνω εταιρεία είναι ελλειψοειδή; Η απάντηση είναι «όχι».

Ας υποθεθεί ότι, άγνωστο στον ερευνητή, το ακριβές ποσοστό των ελλειψοειδών λαμπτήρων είναι 44%. Ένα δείγμα 100 λαμπτήρων μπορεί να περιέχει 42 ελλειψοειδούς. Συνεπώς, τα συμπεράσματα που βασίζονται στην δειγματοληψία πάντοτε υπόκεινται στο στοιχείο *της αβεβαιότητας*.

Η Στατιστική χρησιμοποιεί την ιδέα της *Πιθανότητας* για να μετρήσει την αβεβαιότητα η οποία συνοδεύει μια συμπερασματολογία. Η επιστήμη αυτή μας βοηθά να υπολογίσουμε τις πιθανότητες εκλογής ορισμένων δειγμάτων, κάτω από ορισμένες συνθήκες που αφορούν τον πληθυσμό. Μετά το στάδιο αυτό, ο στατιστικός επιστήμονας χρησιμοποιεί τις πιθανότητες αυτές για να «υπολογίσει» την αβεβαιότητα που συνοδεύει την στατιστική συμπερασματολογία.

Συνεπώς, η μεγάλη συνεισφορά της Στατιστικής Επιστήμης είναι η συναγωγή συμπερασμάτων σε φαινόμενα και πληθυσμούς στα οποία υπάρχει το στοιχείο της αβεβαιότητας.

Επίσης, η πρόβλεψη μελλοντικών παρατηρήσεων, με βάση δεδομένα του παρελθόντος, είναι κύριο έργο της Στατιστικής. Μερικές τεχνικές πρόβλεψης, όπως η απλή Γραμμική Παλινδρόμηση, θα εξετασθούν με λεπτομέρεια στο κεφάλαιο των γραμμικών στατιστικών υποδειγμάτων.

Ανακαιφαλαιώνοντας, παρατηρούμε ότι, η Στατιστική ασχολείται με δύο είδη προβλημάτων: 1) Την οργάνωση και περιγραφή δεδομένων (περιγρα-

φική Στατιστική), (descriptive statistics), και 2) Τη χρήση δειγματικών δεδομένων για τη συναγωγή συμπερασμάτων που αφορούν τους στατιστικούς πληθυσμούς από τους οποίους τα δεδομένα αυτά προέρχονται. (Στατιστική Συμπερασματολογία ή Επαγωγική Στατιστική), (inferential statistics).

Δίνουμε επίσης τους εξής ορισμούς:

### **Ορισμός 1.4.3**

*Παράμετρος* (parameter), είναι ένα αριθμητικό μέγεθος που περιγράφει κάποιο χαρακτηριστικό του στατιστικού πληθυσμού. Οι παράμετροι συμβολίζονται με γράμματα του ελληνικού αλφαβήτου, π.χ. ο μέσος  $\mu$  και η διακύμανση  $\sigma^2$  ενός πληθυσμού.

### **Ορισμός 1.4.4**

Στατιστικό (statistic), είναι ένα αριθμητικό μέγεθος που περιγράφει κάποιο χαρακτηριστικό των δειγματικών παρατηρήσεων. Τα στατιστικά συμβολίζονται με γράμματα του αγγλικού αλφαβήτου, π.χ. ο μέσος  $\bar{x}$  ενός δείγματος ή η διακύμανση  $S^2$  ενός δείγματος.

## **1.5 Έννοια της Στατιστικής Μεταβλητής & οι Διακρίσεις αυτής**

### **Ορισμός 1.5.1**

Μεταβλητή (variable), είναι ένα χαρακτηριστικό των στοιχείων ενός πληθυσμού υπό εξέταση.

Μπορούμε να διακρίνουμε δύο είδη παρατηρήσεων, τις ποιοτικές και τις ποσοτικές.

Οι *ποιοτικές παρατηρήσεις*, αποτελούνται από χαρακτηριστικά, όπως το πολιτικό κόμμα, η θρησκεία, το φύλο κ.λπ.

Οι *ποσοτικές παρατηρήσεις*, αποτελούνται από αριθμούς που αντιπροσωπεύουν μετρήσεις. Επίσης, μπορούμε να διακρίνουμε δύο τύπους ποσοτικών παρατηρήσεων, τις διακριτές και τις συνεχείς.

### **Ορισμός 1.5.2**

Οι διακριτές ποσοτικές παρατηρήσεις (discrete quantitative data) προέρχονται είτε από ένα πεπερασμένο αριθμό πιθανών τιμών, είτε από ένα «μετρήσιμο» αριθμό πιθανών τιμών.

### **Παράδειγμα 1.5.1**

Ας υποθεθεί ότι ρίχνουμε στον αέρα ένα ζευγάρι ζαριών των οποίων η κάθε πλευρά είναι αριθμημένη από το 1 ως το 6. Κάθε φορά που ρίχνουμε τα δυο ζάρια υπολογίζουμε το άθροισμα των αριθμών που είναι επάνω στις πλευρές που φαίνονται. Εάν μετρήσουμε τον αριθμό των ρίψεων, μέχρι να έλθει για πρώτη φορά το άθροισμα 7, μπορούμε να έχουμε κάθε μία από τις τιμές 1,2,3,... Αυτό το σύνολο παρατηρήσεων λέγεται *μετρήσιμο*.

Ο αριθμός των παιδιών μιας οικογένειας μπορεί να είναι 0,1,2,3,4,5,6, ... Αυτές οι παρατηρήσεις είναι πεπερασμένες διακριτές παρατηρήσεις.

### **Ορισμός 1.5.3**

Οι συνεχείς ποσοτικές παρατηρήσεις (continuous quantitative data) προέρχονται από άπειρες πιθανές τιμές (infinitely many values) που μπορούν να παραστήσουν σημεία σε μία συνεχή κλίμακα κατά τέτοιο τρόπο ώστε να μην υπάρχουν κενά.

### Παράδειγμα 1.5.2

Εάν μετρούσαμε τα ποσά (σε λίτρα) ενός ορισμένου τύπου μπύρας σε διάφορα μπουκάλια, θα μπορούσαμε να λάβουμε τιμές όπως 1,026 ή 0,99 ή *κάθε τιμή* μεταξύ του 1,026 και 0,99. Τέτοιες παρατηρήσεις, λέγονται συνεχείς.

Ένα άλλο παράδειγμα είναι το εξής: Τα μεγέθη των παπουτσιών (37,38,39,40, κ.λπ.) είναι διακριτές παρατηρήσεις, ενώ τα πραγματικά μήκη των ποδιών (π.χ. 0,50cm, 0,60cm, 0,55cm κ.λπ.) είναι συνεχείς παρατηρήσεις.

### Ορισμός 1.5.4

Κατηγορικές (nominal) παρατηρήσεις είναι εκείνες που δηλώνουν ονόματα, σύμβολα ή μόνο κατηγορίες. Τα δεδομένα αυτά δεν μπορούν να διαταχθούν σε μία σειρά.

### Παράδειγμα 1.5.3

Η παρατήρηση «ναι, όχι, αναποφάσιτος», η οποία είναι το σύνολο των πιθανών απαντήσεων σε μία ερώτηση δειγματοληπτικής έρευνας, είναι μία κατηγορική παρατήρηση.

### Ορισμός 1.5.5

Οι διαβαθμισμένες ή ποιοτικές (ordinal) παρατηρήσεις είναι εκείνες οι παρατηρήσεις που μπορούν να διαταχθούν σε κάποια σειρά, αλλά οι διαφορές μεταξύ των τιμών των παρατηρήσεων είτε δεν μπορούν να προσδιορισθούν, είτε είναι άνευ σημασίας.



**Παράδειγμα 1.5.4**

Παρατηρήσεις σαν τις ακόλουθες: «καλός», «καλύτερος» ή «κάλιστος» είναι ποιοτικές παρατηρήσεις και δεν μπορούμε να προσδιορίσουμε τις ειδικές διαφορές μεταξύ τέτοιων δεδομένων.