

- 1.1 Ιδιότητες πιθανοτήτων
- 1.2 Μέθοδοι απαρίθμησης
- 1.3 Υπό συνθήκη πιθανότητα
- 1.4 Ανεξάρτητα ενδεχόμενα
- 1.5 Θεώρημα Bayes

## 1.1 ΙΔΙΟΤΗΤΕΣ ΠΙΘΑΝΟΤΗΤΩΝ

Συνήθως είναι δύσκολο να εξηγήσει κανείς στο ευρύ κοινό τι κάνουν οι στατιστικολόγοι, οι επιστήμονες της στατιστικής. Πολλοί μάς θεωρούν «σπασίκες των μαθηματικών» που απολαμβάνουμε να παίζουμε με τους αριθμούς, και πιθανότατα υπάρχει κάποια αλήθεια σε αυτό. Αλλά αν δούμε τη μεγαλύτερη εικόνα, πολλοί αναγνωρίζουν ότι οι στατιστικοί επιστήμονες θα μπορούσαν να είναι εξαιρετικά χρήσιμοι σε πολλές έρευνες.

Σκεφτείτε τα παρακάτω:

- Υπάρχει κάποιο πρόβλημα ή κατάσταση που πρέπει να εξεταστεί και οι στατιστικοί επιστήμονες συχνά καλούνται να συνεργαστούν με αστυνομικούς ερευνητές ή επιστήμονες της έρευνας.
- Έστω ότι απαιτείται κάποιο μέτρο που θα μας βοηθήσει να κατανοήσουμε καλύτερα την κατάσταση. Το πρόβλημα της μέτρησης είναι πολλές φορές εξαιρετικά δύσκολο και η δημιουργία αξιόπιστων μετρικών αποτελεί μια πολύτιμη δεξιότητα. Για παράδειγμα, στην ανώτατη εκπαίδευση, πώς μετράμε την καλή διδασκαλία; Αυτό είναι ένα ερώτημα για το οποίο δεν έχουμε βρει ικανοποιητική απάντηση, αν και στο παρελθόν έχουν χρησιμοποιηθεί διάφορες μετρικές, όπως οι αξιολογήσεις από φοιτητές.
- Αφού αναπτυχθεί ένα όργανο μέτρησης, πρέπει να συλλέξουμε δεδομένα μέσω παρατήρησης – αυτά τα δεδομένα θα μπορούσαν να είναι τα αποτελέσματα ενός ερωτηματολογίου ή ενός πειράματος.
- Χρησιμοποιώντας αυτά τα δεδομένα, οι στατιστικοί επιστήμονες συνοψίζουν τα αποτελέσματα, συχνά με περιγραφική στατιστική και γραφικές μεθόδους.
- Αυτές οι συνόψεις χρησιμοποιούνται τότε για την ανάλυση της κατάστασης. Με αυτόν τον τρόπο, οι στατιστικοί επιστήμονες μπορούν να προχωρήσουν στις ενέργειες που συνολικά αναφέρονται ως στατιστική επαγωγή.
- Τέλος, παρουσιάζεται μια αναφορά, μαζί με κάποιες προτάσεις που βασίζονται στα δεδομένα και στην ανάλυσή τους. Συχνά, μια τέτοια πρόταση θα μπορούσε να είναι η επανάληψη του ερωτηματολογίου ή του πειράματος, ίσως αλλάζοντας κάποιες από τις ερωτήσεις ή τους παράγοντες που περιλαμβάνονται. Με αυτόν τον τρόπο χρησιμοποιείται η στατιστική σε αυτό που αναφέρεται ως επιστημονική μέθοδος, επειδή συχνά η ανάλυση των δεδομένων προτείνει άλλα πειράματα. Αναλόγως, ο επιστήμονας οφείλει να εξετάσει διαφορετικούς δρόμους στην αναζήτησή του για μια απάντηση και, ως εκ τούτου, να εκτελέσει παρόμοια πειράματα ξανά και ξανά.

Ο κλάδος της στατιστικής ασχολείται με τη *συλλογή και ανάλυση δεδομένων*. Όταν γίνονται μετρήσεις, ακόμα και φαινομενικά υπό τις ίδιες συνθήκες, τα αποτελέσματα συνήθως διαφέρουν. Παρά τη μεταβλητότητα, ένας στατιστικός επιστήμονας προσπαθεί να βρει ένα μοτίβο. Ωστόσο, λόγω του «θορύβου», δεν προσαρμόζονται όλα τα δεδομένα στο μοτίβο. Παρά τη μεταβλητότητα, ο στατιστικός επιστήμονας είναι υποχρεωμένος να προσδιορίσει ποιος είναι ο καλύτερος τρόπος για να περιγράψει το μοτίβο. Ως εκ τούτου, οι στατιστικοί επιστήμονες γνωρίζουν ότι θα γίνουν λάθη στην ανάλυση δεδομένων και προσπαθούν να τα ελαχιστοποιήσουν όσο το δυνατόν περισσότερο και στη συνέχεια να ορίσουν τα όρια για τα πιθανά σφάλματα. Λαμβάνοντας υπόψη αυτά τα όρια, οι υπεύθυνοι για τη λήψη αποφάσεων μπορούν να αποφασίσουν πόση εμπιστοσύνη θέλουν να δείξουν στα δεδομένα και στην ανάλυσή τους. Αν τα όρια παρέχουν μεγάλα περιθώρια, ίσως πρέπει να συλλεχθούν περισσότερα δεδομένα. Αν όμως τα όρια είναι στενά, ο υπεύθυνος της μελέτης ενδεχομένως να πρέπει να λάβει μια απόφαση και να προχωρήσει αναλόγως.

Η μεταβλητότητα είναι γεγονός και οι κατάλληλες στατιστικές μέθοδοι μπορούν να μας βοηθήσουν να κατανόησουμε τα δεδομένα που συλλέγονται σε συνθήκες εγγενούς μεταβλητότητας. Λόγω αυτής της μεταβλητότητας, πολλές αποφάσεις που λαμβάνονται περιλαμβάνουν αβεβαιότητες. Για παράδειγμα, ένας ιατρικός ερευνητής ενδιαφέρεται για την αποτελεσματικότητα ενός νέου εμβολίου για την παρωτίτιδα, ένας γεωπόνος πρέπει να αποφασίσει αν μια αύξηση της απόδοσης μπορεί να αποδοθεί σε ένα νέο στέλεχος σιταριού, ένας μετεωρολόγος προσπαθεί να προβλέψει την πιθανότητα βροχής, ο νομοθέτης πρέπει να αποφασίσει αν η μείωση των ορίων ταχύτητας θα μπορούσε να οδηγήσει σε λιγότερα ατυχήματα, ο υπεύθυνος της γραμματείας ενός πανεπιστημίου πρέπει να προβλέψει τις επιδόσεις ενός πρωτοετούς φοιτητή στο πανεπιστήμιο, ένας βιολόγος ενδιαφέρεται να εκτιμήσει το μέγεθος των νυχιών για ένα συγκεκριμένο είδος πτηνού, ένας οικονομολόγος επιθυμεί να εκτιμήσει το ποσοστό ανεργίας και ένας περιβαλλοντολόγος προσπαθεί να διαπιστώσει αν οι νέοι έλεγχοι οδήγησαν στη μείωση της ρύπανσης.

Κατά την ανασκόπηση του προηγούμενου (σχετικά σύντομου) καταλόγου πιθανών τομέων εφαρμογής της στατιστικής, ο αναγνώστης πρέπει να αναγνωρίσει ότι η καλή στατιστική συνδέεται στενά με την προσεκτική σκέψη σε πολλές έρευνες. Για παράδειγμα, οι φοιτητές πρέπει να εκτιμούν τη χρήση της στατιστικής στον ατελείωτο κύκλο της επιστημονικής μεθόδου. Παρατηρούμε τη φύση και θέτουμε ερωτήματα, εκτελούμε πειράματα και συλλέγουμε δεδομένα που ρίχνουν φως σε αυτά τα ερωτήματα, αναλύουμε τα δεδομένα και συγκρίνουμε τα αποτελέσματα της ανάλυσης με ό,τι έχουμε σκεφτεί προηγουμένως, θέτουμε νέα ερωτήματα κ.ο.κ. Ή, αν προτιμάτε, η στατιστική είναι σαφώς ένα κομμάτι του σημαντικού κύκλου «σχεδιασμός-εφαρμογή-μελέτη-ενέργεια»: Τίθενται ερωτήματα και έρευνες μελετώνται και εκτελούνται. Τα δεδομένα που προκύπτουν μελετώνται, αναλύονται και επί αυτών εκτελούνται εργασίες, με αποτέλεσμα να εγείρονται συχνά νέα ερωτήματα.

Υπάρχουν πολλές πτυχές της στατιστικής επιστήμης. Κάποιοι άνθρωποι μπορεί να ενδιαφέρονται για κάποιο θέμα και να συλλέγουν δεδομένα σε μια προσπάθεια να εξαγάγουν νόημα από τις παρατηρήσεις τους. Σε ορισμένες περιπτώσεις, οι απαντήσεις είναι προφανείς και δεν απαιτούνται ιδιαίτερες γνώσεις των στατιστικών μεθόδων. Αλλά αν κάποιος συμμετέχει σε βάθος σε πολλές έρευνες, σύντομα συνειδητοποιεί ότι είναι αναγκαίο ένα θεωρητικό υπόβαθρο ώστε να μπορεί να κατανοήσει καλύτερα τη δομή του σφάλματος που σχετίζεται με τις διάφορες εκτιμήσεις των προτύπων. Με άλλα λόγια, σε κάποιο σημείο θα απαιτηθούν τα κατάλληλα μοντέλα πιθανοτήτων και μαθηματικών, ώστε να είναι δυνατή η κατανόηση περίπλοκων συνόλων δεδομένων. Η στατιστική και τα πιθανοτικά ή πιθανοθεωρητικά θεμέλια στα οποία βασίζονται οι στατιστικές μέθοδοι μπορούν να παρέχουν τα μοντέλα που θα βοηθήσουν τους ανθρώπους στην κατανόηση των παραπάνω. Επομένως, σε αυτό το βιβλίο ενδιαφερόμαστε περισσότερο για τις μαθηματικές και όχι για τις εφαρμοσμένες πτυχές της στατιστικής. Σε κάθε περίπτωση, όμως, παρέχουμε αρκετά παραδείγματα που βασίζονται σε πραγματικές καταστάσεις και θα βοηθήσουν τον αναγνώστη να αντιληφθεί πολλές σημαντικές εφαρμογές των μεθόδων της στατιστικής.

Κατά τη μελέτη της στατιστικής, εξετάζουμε πειράματα για τα οποία το αποτέλεσμα δεν μπορεί να προβλεφθεί με βεβαιότητα. Τέτοια πειράματα ονομάζονται **τυχαία πειράματα** (random experiments). Αν και το συγκεκριμένο αποτέλεσμα ενός τυχαίου πειράματος δεν μπορεί να προβλεφθεί με βεβαιότητα προτού εκτελεστεί το πείραμα, το σύνολο όλων των πιθανών αποτελεσμάτων είναι γνωστό και μπορεί να περιγραφεί, ίσως και να απαριθμηθεί. Το σύνολο όλων των πιθανών αποτελεσμάτων συμβολίζεται με  $S$  και ονομάζεται **δειγματοχώρος**

ή **δειγματικός χώρος** (sample space). Δεδομένου ενός δειγματοχώρου  $S$ , έστω ότι το  $A$  είναι μέρος ενός συνόλου αποτελεσμάτων στο  $S$ , δηλαδή,  $A \subset S$ . Τότε, το  $A$  ονομάζεται **ενδεχόμενο** (event). Όταν το τυχαίο πείραμα εκτελείται και το αποτέλεσμα του πειράματος βρίσκεται στο  $A$ , λέμε ότι **προέκυψε** (ή έλαβε χώρα) το **ενδεχόμενο**  $A$ .

Κατά τη μελέτη των πιθανοτήτων, οι λέξεις **σύνολο** (set) και **ενδεχόμενο** είναι συνώνυμες και θα προτείνουμε στον αναγνώστη να μελετήσει την **άλγεβρα των συνόλων** (algebra of sets). Εν προκειμένω, θα θυμίσουμε στον αναγνώστη κάποιους όρους:

- το  $\emptyset$  συμβολίζει το **κενό** (null ή empty) σύνολο.
- το  $A \subset B$  σημαίνει ότι το  $A$  είναι **υποσύνολο** (subset) του  $B$ .
- το  $A \cup B$  είναι η **ένωση** (union) των  $A$  και  $B$ .
- το  $A \cap B$  είναι η **τομή** (intersection) των  $A$  και  $B$ .
- το  $A'$  είναι το **συμπλήρωμα** (complement) του  $A$  (δηλαδή, όλα τα στοιχεία στο  $S$  που δεν ανήκουν στο  $A$ ).

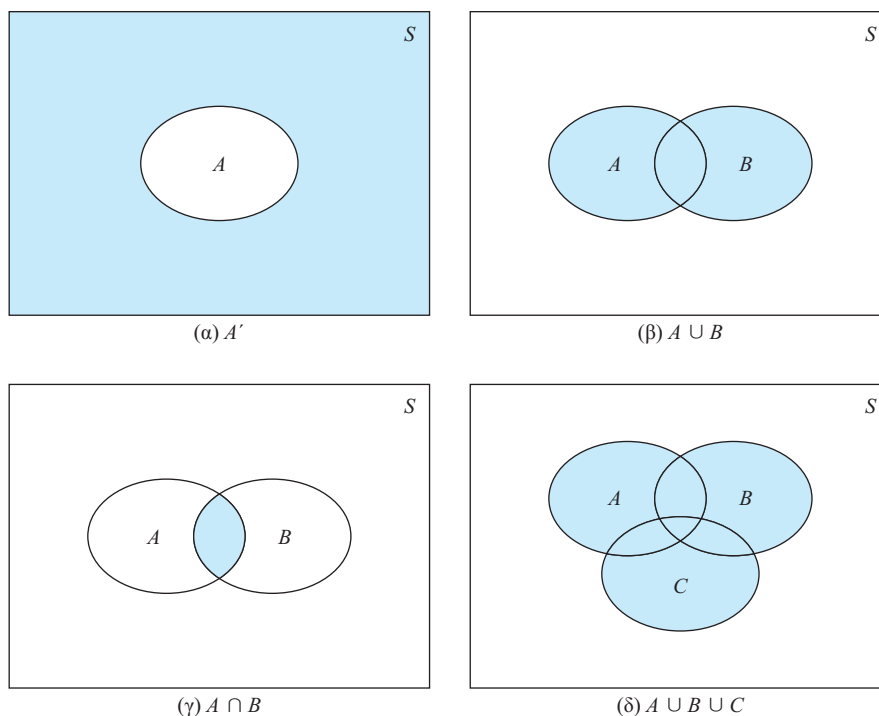
Κάποια από αυτά τα σύνολα απεικονίζονται από τις σκιασμένες περιοχές στο Σχήμα 1.1-1, όπου το  $S$  είναι το εσωτερικό των ορθογωνίων. Τέτοια σχήματα ονομάζονται **διαγράμματα Venn** (Venn diagrams).

Μερικοί από τους ειδικούς όρους που έχουν σχέση με τα ενδεχόμενα και χρησιμοποιούνται συχνά από στατιστικούς επιστήμονες είναι οι εξής:

1. Τα  $A_1, A_2, \dots, A_k$  είναι **αμοιβαία αποκλειόμενα ενδεχόμενα** (mutually exclusive events), που σημαίνει ότι  $A_i \cap A_j = \emptyset, i \neq j$ . δηλαδή τα  $A_1, A_2, \dots, A_k$  είναι ασυμβίβαστα (ή ξένα) σύνολα.
2. Τα  $A_1, A_2, \dots, A_k$  είναι **εξαντλητικά (ή πλήρη) ενδεχόμενα** (exhaustive events), που σημαίνει ότι  $A_1 \cup A_2 \cup \dots \cup A_k = S$ .

Επομένως, αν τα  $A_1, A_2, \dots, A_k$  είναι **αμοιβαία αποκλειόμενα και εξαντλητικά** ενδεχόμενα, γνωρίζουμε ότι  $A_i \cap A_j = \emptyset, i \neq j$  και  $A_1 \cup A_2 \cup \dots \cup A_k = S$ .

Οι πράξεις συνόλων ικανοποιούν αρκετές ιδιότητες. Για παράδειγμα, αν τα  $A, B$  και  $C$  είναι υποσύνολα του  $S$ , έχουμε τα εξής:



Σχήμα 1.1-1 Άλγεβρα συνόλων