

1. Βασικές έννοιες της Στατιστικής

1.1. Η στατιστική μέθοδος

1.1.1. Η στατιστική μέθοδος χρησιμοποιείται για την εξαγωγή συμπερασμάτων από παρατηρούμενα δεδομένα τα οποία αποτελούν μετρήσεις που προέρχονται είτε από την παρατήρηση, είτε από κάποιο πείραμα. Σκοπός της Στατιστικής είναι να περιγράψει τα διάφορα φαινόμενα – αστικά, κοινωνικά, οικονομικά, βιολογικά, δημογραφικά κ.λπ. – στηριζόμενη σε κάποιο Μαθηματικό μοντέλο. Το μοντέλο αυτό μπορεί να είναι είτε καθοριστικό (Deterministic), είτε στοχαστικό (stochastic). Τα καθοριστικά μοντέλα στηρίζονται σε κάποια μαθηματική σχέση και με βάση αυτή τη σχέση περιγράφουν το φαινόμενο. Τα στοχαστικά μοντέλα περιγράφουν καταστάσεις που δεν μπορούν να καθοριστούν με κάποια μαθηματική σχέση, δεν περιγράφονται δηλαδή με κάποιο καθοριστικό μοντέλο. Παραδείγματα στοχαστικών μοντέλων υπάρχουν πολλά: οι διάφορες μεταβολές σε αστικά φαινόμενα, κυκλοφοριολογικές μεταβολές όπως π.χ. απ' τη δημιουργία μιας νέας αρτηρίας, μεταβολές στις σχέσεις παραγωγής και εργασίας όπως π.χ. με τη δημιουργία ενός εργοστασίου σε κάποιον τομέα ενός αστικού κέντρου, ή μεταβολές σε οικονομικά δεδομένα με τη δημιουργία ενός εμπορικού κέντρου. Ακόμα, οι καιρικές μεταβολές, η απόδοση, μιας επιχείρησης, η επίδραση μιας νέας ασθένειας στον πληθυσμό, ολ' αυτά αποτελούν στοχαστικά μοντέλα τα οποία περιγράφουν το φαινόμενο με βάση το νόμο των πιθανοτήτων, την πιθανοθεωρητική δηλαδή συμπεριφορά του αποτελέσματος.

Η στατιστική λοιπόν μέθοδος, διαφέρει απ' την κατηγορική ή παραγωγική μέθοδο (deductive method), που χρησιμοποιείται κύρια στα καθαρά Μαθηματικά.

Για να δούμε αυτή τη διαφορά, ας πάρουμε ένα παράδειγμα παραγωγικού συλλογισμού. Αν είναι:

(α) $A = B$ και (β) $B = \Gamma$, τότε είναι (γ) $A = \Gamma$

Στον κατηγορικό δηλαδή συλλογισμό, είναι απόλυτα βέβαιη η πρόταση (γ), όταν ισχύουν οι προτάσεις (α) και (β).

Η κατάσταση στη Στατιστική είναι διαφορετική και οι συλλογισμοί της είναι κύρια «επαγωγικοί» (inductive).

1.1.2. Ας πάρουμε ένα παράδειγμα στο οποίο θα φανεί η στατιστική μέθοδος.

Ας υποθέσουμε ότι σχηματίζουμε ένα «αντιπροσωπευτικό» δείγμα (τους ορισμούς των εννοιών αυτών αυτών θα τους δούμε στις επόμενες παραγράφους) από τους εργαζόμενους σε μια αστική περιοχή και διαπιστώνουμε ότι:

α) το 20% των εργαζόμενων του δείγματος εργάζονται στη βιομηχανία.

Από αυτή την πρόταση θα μπορούσαμε να βγάλουμε, με τον επαγωγικό συλλογισμό, το συμπέρασμα:

β) το 20% όλων των εργαζόμενων στην αστική περιοχή εργάζονται στη βιομηχανία.

Καταλαβαίνουμε όμως ότι το συμπέρασμα αυτό δεν είναι απόλυτα βέβαιο και είναι ανάγκη να το περιορίσουμε, να το δούμε με μία επιφύλαξη, ότι δηλαδή το συμπέρασμα είναι «περίπου σωστό» ή «κατά προσέγγιση σωστό» και τούτο γιατί δε ρωτήθηκαν όλοι οι εργαζόμενοι στην αστική περιοχή. Η πληροφορία που έχουμε από το δείγμα μας επιτρέπει να συμπεράνουμε ότι οι εργαζόμενοι στη βιομηχανία είναι περίπου 20% του συνόλου όλων των εργαζόμενων. Αυτό το συμπέρασμα είναι το αποτέλεσμα ενός επαγωγικού συλλογισμού.

Ένα στατιστικό συμπέρασμα είναι μια μορφή επαγωγικού συμπεράσματος, που επιτρέπει στο μελετητή να είναι σχετικά ακριβέστερος στην αοριστία «περίπου σωστό» ή «κατά προσέγγιση σωστό», διασαφηνίζοντας ότι, με κίνδυνο να γίνει σφάλμα π.χ. 5% (δηλαδή με πιθανότητα 0,95), το ποσοστό των εργαζόμενων στη βιομηχανία βρίσκεται μεταξύ των ποσοστών 14% και 26% του συνόλου όλων των εργαζόμενων.

Οι στατιστικές μέθοδοι είναι ακριβώς οι τεχνικές για να εξαχθούν στατιστικά συμπεράσματα και δίνουν διαδικασίες για τον υπολογισμό τέτοιων ορίων με πιθανότητες καθώς και για τον έλεγχο διαφόρων στατιστικών υποθέσεων.

1.1.3. Η Στατιστική είναι ένας κλάδος των εφαρμοσμένων μαθηματικών, που, χρησιμοποιώντας δείγματα από πληθυσμούς και ακολουθώντας

στατιστικές μεθόδους, βγάζει συμπεράσματα που αναφέρονται στους πληθυσμούς από τους οποίους προήλθαν τα δείγματα.

1.2. Βασικές έννοιες και ορισμοί

1.2.1. *Στατιστικό σύνολο* λέγεται κάθε σύνολο οποιωνδήποτε «πραγμάτων», που θέλουμε να μελετήσουμε με στατιστικές μεθόδους. Κάθε στοιχείο του στατιστικού συνόλου καλείται *στατιστικό στοιχείο* ή απλούστερα *στοιχείο* ή *άτομο* ή *μονάδα*.

Διακρίνουμε δύο τύπους στατιστικών συνόλων: τον πληθυσμό και το δείγμα.

Πληθυσμός (και αν ακόμη δεν πρόκειται για ανθρώπους) είναι το σύνολο όλων των στατιστικών ατόμων. Αν θέλουμε π.χ. να μελετήσουμε τα νοικοκυριά μιας περιοχής, πληθυσμός είναι το σύνολο των νοικοκυριών αυτών και όχι το σύνολο των ανθρώπων που τα αποτελούν. Το σύνολο όλων των κατοίκων της περιοχής είναι άλλος πληθυσμός, διαφορετικός από τον πληθυσμό που αποτελούν τα νοικοκυριά.

Κάθε υποσύνολο ενός πληθυσμού καλείται *δείγμα* του πληθυσμού αυτού.

Το πλήθος των στοιχείων ενός στατιστικού συνόλου λέγεται *μέγεθος του στατιστικού συνόλου*. Έτσι έχουμε μέγεθος πληθυσμού και μέγεθος δείγματος. Συνήθως εξετάζουμε *πεπερασμένους πληθυσμούς*, δηλαδή πληθυσμούς που έχουν πεπερασμένο πλήθος στοιχεία.

1.2.2. Θέλουμε να μελετήσουμε ορισμένα *χαρακτηριστικά* ενός στατιστικού συνόλου. Π.χ. στον πληθυσμό των νοικοκυριών κάποιας περιοχής, ένα χαρακτηριστικό είναι το «πλήθος των μελών των νοικοκυριών», ένα άλλο είναι «η καταγωγή» τους, ή το «μηνιαίο τους εισόδημα» κ.τ.λ. Τα χαρακτηριστικά λέγονται πολλές φορές και *μεταβλητές*.

Τα χαρακτηριστικά χωρίζονται σε δύο κατηγορίες: τα ποιοτικά και τα ποσοτικά, που θα τα εξετάσουμε παρακάτω.

1.2.3. *Ποιοτικό* λέγεται το χαρακτηριστικό που δε μεταβάλλεται με συνεχή τρόπο, δεν είναι επιδεκτικό μέτρησης, αλλά είναι δυνατό μόνο να ταξινομηθεί. Ποιοτικά χαρακτηριστικά είναι π.χ. το φύλο των ανθρώπων (αρσενικό, θηλυκό), η οικογενειακή κατάσταση (άγαμος, έγγαμος, σε χηρεία, διαζευγμένος), η καταγωγή (ντόπιος, πρόσφυγας, εσωτερικός μετανάστης) κ.τ.λ. Με ένα ποιοτικό χαρακτηριστικό, το στατιστικό σύνολο χωρίζεται σε *τάξεις*. Έτσι, αν το χαρακτηριστικό είναι η καταγωγή τα νοικοκυριά κάποιας περιοχής μπορούν να χωριστούν σε τρεις τάξεις: στην τάξη των προσφυγικών νοικοκυριών, των ντόπιων νοικοκυριών και των νοικοκυριών

των εσωτερικών μετανάστών. Πολλές φορές, για απλότητα, στις τάξεις δίνουμε έναν κωδικό αριθμό. Το πλήθος των στοιχείων κάθε τάξης λέγεται *μέγεθος της τάξης*. Είναι ουσιαστικό να τονίσουμε ότι οι τάξεις, στις οποίες χωρίστηκε ένα στατιστικό σύνολο, είναι σύνολα *ξένα μεταξύ τους ανά δύο*.

Όταν έχουμε κάνει το διαχωρισμό του στατιστικού συνόλου σε τάξεις με κάποιο χαρακτηριστικό, τότε λέμε ότι έχουμε κάνει μία *κατανομή του στατιστικού συνόλου* ως προς το θεωρούμενο χαρακτηριστικό.

1.2.4. Ας υποθέσουμε ότι έχουμε ένα στατιστικό σύνολο E (πληθυσμό ή δείγμα) που το μέγεθός του είναι n . Με ένα ποιοτικό χαρακτηριστικό, το E χωρίζεται σε k τάξεις E_1, E_2, \dots, E_k , που τα μεγέθη τους ας υποθέσουμε ότι είναι αντίστοιχα n_1, n_2, \dots, n_k . Έχουμε έτσι:

$$E_1 \cup E_2 \cup \dots \cup E_k = E, \quad (E_i \cap E_j = \emptyset \cdot i, j = 1, 2, \dots, k, i \neq j), \quad (1.2.1)$$

$$n_1 + n_2 + \dots + n_k = n \quad (1.2.2)$$

Συχνότητα f_i της τάξης E_i καλείται ο αριθμός:

$$f_i = \frac{n_i}{n}, \quad (1.2.3)$$

και είναι πάντα

$$0 \leq f_i \leq 1 \quad (i = 1, 2, \dots, k) \quad (1.2.4)$$

Επίσης από τη σχέση (1.2.2) προκύπτει αμέσως ότι

$$f_1 + f_2 + \dots + f_k = 1 \quad (1.2.5)$$

Πολλές φορές, αντί της συχνότητας f_i μιας τάξης, παίρνουμε το *ποσοστό στα εκατό* της τάξης, που είναι ο αριθμός:

$$100 f_i \quad (i = 1, 2, \dots, k)$$

Αν τοποθετήσουμε τις τάξεις E_i σε μία ορισμένη σειρά, π.χ. στη σειρά E_1, E_2, \dots, E_n , τότε *αθροιστική συχνότητα της τάξης* E_i καλείται το άθροισμα τη συχνότητας της τάξης E_i και των συχνοτήτων των τάξεων που προηγούνται της E_i δηλαδή η αθροιστική συχνότητα της E_i είναι ο αριθμός:

$$f_1 + f_2 + \dots + f_i = \frac{n_1 + n_2 + \dots + n_i}{n} \quad (1.2.6)$$

Έτσι η αθροιστική συχνότητα της τάξης E_1 είναι f_1 , της τάξης E_2 είναι $f_1 + f_2$, κ.ο.κ., τέλος της τάξης E_k είναι $f_1 + f_2 + \dots + f_k = 1$.

Είναι φανερό ότι, αν αλλάξουμε τη σειρά που τοποθετούμε τις τάξεις, η αθροιστική συχνότητα κάθε τάξης μεταβάλλεται.