

Στατιστική στις Γεωπονικές Επιστήμες

1 Βασικές Έννοιες στη Στατιστική

Εισαγωγή

Η Στατιστική είναι κλάδος των Εφαρμοσμένων Μαθηματικών ο οποίος ασχολείται με τη μεθοδική συλλογή, παρουσίαση, ανάλυση, περιγραφή και ερμηνεία στατιστικών στοιχείων, με απώτερο στόχο την εξαγωγή ασφαλών συμπερασμάτων για λήψη ορθών αποφάσεων. Η επιστήμη της Στατιστικής επιχειρεί να εξαγάγει έγκυρη γνώση χρησιμοποιώντας εμπειρικά δεδομένα παρατήρησης ή και πειράματος.

Οι μέθοδοι έρευνας στη Στατιστική δε διαφέρουν εκείνων που χρησιμοποιούνται στα Μαθηματικά. Μας βοηθούν στην επεξεργασία των στατιστικών στοιχείων κατά τέτοιο τρόπο ώστε να αποκαλύπτουν τι κρύβεται πίσω από τα στοιχεία/δεδομένα. Οι στατιστικές μέθοδοι είναι μεν αντικειμενικοί και ακριβείς, απαιτείται όμως μεγάλη προσοχή από τους ερευνητές στην εφαρμογή και την ερμηνεία των στατιστικών δεδομένων. Η ακρίβεια των αποτελεσμάτων των στατιστικών μεθόδων εξαρτάται από την ακρίβεια και αντικειμενικότητα των στοιχείων που έχουν συλλεγεί.

Η Στατιστική χωρίζεται σε δύο μεγάλους τομείς:

1. **την Περιγραφική Στατιστική:** Η περιγραφική στατιστική αποτελεί ένα σύνολο μεθόδων και εργαλείων (γραφικών και ποσοτικών) για την αποτελεσματική παρουσίαση δεδομένων μέσω της σύνοψης, της ομαδοποίησης και της απεικόνισής τους. Τα αποτελέσματα ενός πειράματος συνήθως δημιουργούν ένα μεγάλο αριθμό δεδομένων. Κρίνεται λοιπόν αναγκαία η εύρεση δι-αδικασιών, με τις οποίες τα αποτελέσματα αυτά μπορούν να οργανωθούν και να παρουσιαστούν με απλό και εύληπτο τρόπο. Αυτός ακριβώς είναι και ο στόχος των περιγραφικών στατιστικών δεικτών, δηλαδή να παρέχουν μεθόδους που απλοποιούν και διευκολύνουν την οργάνωση και παρουσίαση των αποτελεσμάτων. Στόχος είναι η συνοπτική περιγραφή-παρουσίαση ενός τυχαίου δείγματος, ώστε αφενός να αναδεικνύονται τα σημαντικά χαρακτηριστικά του και αφετέρου να αποδίδεται μια πρώτη εικόνα του πληθυσμού. Φυσικά, τονίζεται ότι τα οποιαδήποτε συμπεράσματα που απορρέουν από την περιγραφική ανάλυση ισχύουν μόνο για το συγκεκριμένο δείγμα και δεν επεκτείνονται στον πληθυσμό χωρίς περαιτέρω ανάλυση. Με την περιγραφική στατιστική περιγράφουμε τις μεθόδους που ασχολούνται με τη συλλογή, παρουσίαση και χαρακτηρισμό (ταξινόμηση) των δεδομένων ανάλογα με το είδος των χαρακτηριστικών που περιγράφουν (μετρούν).
1. **την Επαγωγική Στατιστική:** Κατά την επαγωγική στατιστική αναπτύσσονται στατιστικές μεθοδολογίες εκμετάλλευσης της πληροφορίας που παρέχεται από ένα τυχαίο δείγμα για την εξαγωγή συμπερασμάτων για τον (άγνωστο) πληθυσμό, με συγκεκριμένη βεβαιότητα. Ο στόχος είναι οι εκτιμήσεις άγνωστων ποσοτήτων (ή παραμέτρων) ενός ή δύο (διακριτών ή συνεχών) πληθυσμών και οι διάφοροι έλεγχοι υποθέσεων, με έμφαση στις ερμηνείες των συμπερασμάτων. Η

στατιστική συμπερασματολογία παρουσιάζει θεωρητικά συνεπείς τρόπους γενίκευσης των στατιστικών συμπερασμάτων για το σύνολο του πληθυσμού από τον οποίο προέρχεται το τυχαίο δείγμα της ανάλυσης (διαστήματα εμπιστοσύνης και στατιστικοί έλεγχοι υποθέσεων). Ορίζονται οι μέθοδοι που μας βοηθούν να εκτιμήσουμε τα χαρακτηριστικά ενός πληθυσμού με τα αποτελέσματα που προκύπτουν από τις παρατηρήσεις ενός δείγματος.

Πληθυσμός - Δείγμα

Ένα σύνολο του οποίου τα στοιχεία μελετάμε ως προς κάποιο χαρακτηριστικό τους, λέγεται **πληθυσμός**. Εδώ πρέπει να γίνει μια διάκριση μεταξύ **πεπερασμένου** (finite) και **μη πεπερασμένου ή άπειρου** (infinite) πληθυσμού. π.χ. οι μαθητές οι οποίοι φοιτούν στη πρώτη τάξη των σχολείων Μέσης Εκπαίδευσης είναι πεπερασμένος πληθυσμός, διότι μπορούμε να γνωρίζουμε ακριβώς πόσοι είναι οι μαθητές αριθμητικώς. Το πλήθος των ρίψεων ενός νομίσματος μπορεί να θεωρηθεί ότι είναι πληθυσμός άπειρος, διότι μπορούμε θεωρητικά να αυξάνουμε τον αριθμό των ρίψεων απεριόριστα.

Σε πολλές περιπτώσεις ο στατιστικολόγος εξαιτίας του μεγάλου αριθμού των στοιχείων του συνόλου, θεωρεί ως μη πεπερασμένο τον πληθυσμό, μολονότι είναι πεπερασμένος. π.χ. ο πληθυσμός της Ευρώπης είναι τόσο μεγάλος, ώστε στατιστικά να θεωρείται ότι είναι άπειρος.

Επειδή συνήθως ο πληθυσμός είναι μεγάλου μεγέθους και πρακτικά είναι δύσκολο, εάν όχι αδύνατο, να ληφθούν όλες οι μετρήσεις, γίνεται εκλογή/επιλογή ενός υποσυνόλου του πληθυσμού, το οποίο καλούμε **δείγμα** (sample) και το οποίο οπωσδήποτε πρέπει να επιλεγεί κατά τέτοιο τρόπο ώστε να αντιπροσωπεύει όσο το δυνατόν καλύτερα το πληθυσμό στο σύνολό του. Στηριζόμενος σε αυτό το δείγμα ο ερευνητής, χρησιμοποιώντας κατάλληλες μεθόδους, καταλήγει σε συμπεράσματα τα οποία δέχεται ότι ισχύουν και για το πληθυσμό από τον οποίο προήλθε το δείγμα. Τα συμπεράσματα τα οποία προκύπτουν από το δείγμα, τότε μόνο θα ισχύουν για το πληθυσμό, όταν το δείγμα είναι τυχαίο, ανεξάρτητο και αντιπροσωπευτικό. Το δείγμα είναι τυχαίο, όταν η εκλογή γίνεται κατά τέτοιο τρόπο, ώστε κάθε στοιχείο του πληθυσμού να έχει την ίδια πιθανότητα να (συμ)περιληφθεί σε αυτό. Ανεξάρτητο είναι, όταν η εκλογή ενός δεύτερου δείγματος από τον πληθυσμό δεν εξαρτάται από την εκλογή του πρώτου δείγματος. Η αντιπροσωπευτικότητα ενός δείγματος καθορίζεται από το μέγεθος του τυχαίου δείγματος. Όσο μεγαλύτερο είναι το δείγμα, τόσο πιο αξιόπιστα είναι τα συμπεράσματα που εξαγονται από αυτό.

Μεταβλητή

Το χαρακτηριστικό (π.χ. η ομάδα προτίμησης στο ποδόσφαιρο) ως προς το οποίο μελετάμε τα στοιχεία ενός πληθυσμού, ονομάζεται **μεταβλητή** (variable). Συνήθως οι μεταβλητές εκφράζονται ως αριθμητικές τιμές. π.χ. το βάρος, το ύψος, η ηλικία, εφόσον μπορούν να μετρηθούν και να εκφραστούν με διαφορετικές τιμές, είναι μεταβλητές.

Η μεταβλητή μπορεί να είναι συνεχής ή ασυνεχής. **Συνεχής** καλείται η μεταβλητή η οποία μπορεί να λάβει, θεωρητικώς τουλάχιστον όλες τις τιμές σε ένα διάστημα, δηλ. κινείται μεταξύ μιας ελάχιστης και μιας μέγιστης τιμής. Το ύψος για παράδειγμα είναι μια συνεχής μεταβλητή. Μια μεταβλητή είναι **ασυνεχής** (discrete), όταν λαμβάνει τιμές μόνο φυσικούς αριθμούς. Ο αριθμός των μαθητών σε μία τάξη είναι ασυνεχής μεταβλητή, διότι λαμβάνει ως τιμές μόνο φυσικούς αριθμούς.

Οι περισσότερες μεταβλητές είναι συνεχείς. Συνήθως όμως τις συνεχείς μεταβλητές τις μεταχειριζόμαστε ως ασυνεχείς, αφού τις καταγράφουμε στο πλησιέστερο ακέραιο αριθμό. Οι καταγραφές

γίνονται βάσει ορισμένων κανόνων και εντός προκαθορισμένων ορίων. Έτσι ο βαθμός 15,7 καταγράφεται στο πλησιέστερο ακέραιο, δηλ. το 16. ο βαθμός 16 αντιπροσωπεύει όλους τους βαθμούς, ο οποίοι ευρίσκονται μεταξύ 15,5 και 16,5⁻ (16, 4999...). Το πλην (-) στο 16,5 έχει την έννοια ότι δεν περιλαμβάνεται στο 16, ενώ το 15,5 περιλαμβάνεται. Το 16,5 περιλαμβάνεται σε διαφορετικό διάστημα που αντιπροσωπεύεται από το 17. Με το τρόπο αυτό κανένας βαθμός δε γράφεται δύο φορές σε διαδοχικά διαστήματα. Με το τρόπο αυτό οι βαθμοί εκπροσωπούν μια συνεχή μεταβλητή.

Δειγματική κύμανση

Δειγματική κύμανση (range) καλείται η διαφορά της μικρότερης από τη μεγαλύτερη των τιμών που έχει λάβει μια μεταβλητή, αυξημένη κατά μία μονάδα. Η δειγματική κύμανση δείχνει το εύρος της διασποράς των τιμών μιας μεταβλητής π.χ. σε ένα διαγώνισμα μιας τάξης έχουμε τους βαθμούς 4, 15, 16, 16, ... , 18, 19, 19, 20, η δειγματική κύμανση είναι $(20 - 4) + 1 = 17$ μολονότι μόνο ένας βαθμός υπάρχει κάτω του 15.

Διαστήματα - Τάξεις

Πολλές φορές τα στοιχεία του υπό εξέταση συνόλου που έχουν συλλεγεί είναι αναγκαίο να τοποθετηθούν κατά τρόπο τέτοιο ώστε εκ πρώτης όψεως να μπορούμε να εξάγουμε στοιχειώδη συμπεράσματα και αν είναι δυνατόν να ανακαλύψουμε παράγοντες που επιδρούν στα στοιχεία του συνόλου. Για το σκοπό αυτό ομαδοποιούμε τα στοιχεία του συνόλου δηλ. το ποθετούμε τα στοιχεία σε μία μικρή ομάδα σύμφωνα με κάποιο κανόνα. Το σύνολο των μικρών ομάδων, εντός των οποίων τοποθετούμε τις παρατηρήσεις ή μετρήσεις μας, καλούνται **τάξεις**. Το εύρος κάθε τάξης καλείται **διάστημα τάξεως** ή **διάστημα** (class interval). π.χ. το σύμβολο 75-79 είναι ένα διάστημα τάξης και περιλαμβάνει τα στοιχεία του συνόλου με αριθμούς 75, 76, 77, 78, 79. Οι αριθμοί 75 και 79 ονομάζονται κατώτερο και ανώτερο όριο, αντίστοιχα. Κατά την ομαδοποίηση των στοιχείων δεν υπάρχουν αυστηροί κανόνες οι οποίοι μας ορίζουν το τρόπο ευρέσεως του εύρους του διαστήματος και του πλήθους των τάξεων.