

Σημειώσεις για το μάθημα

ΣΤΑΤΙΣΤΙΚΗ Ι

Θεοδόσης Δημητράκος

E-mail: dimitheo@aegean.gr

Μάθημα 1

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ ΣΤΗ ΣΤΑΤΙΣΤΙΚΗ

§1.1 Τι είναι η Στατιστική;

Ο όρος Στατιστική είναι μάλλον οικείος στους περισσότερους. Οι εφημερίδες και τα ραδιόφωνα συχνά αναφέρονται στα τροχαία ατυχήματα, στις στατιστικές των γεννήσεων, στις τιμές των μετοχών του Χρηματιστηρίου Αθηνών, στις θερμοκρασίες της Ε.Μ.Υ. κλπ. Υπό αυτή την έννοια ο όρος Στατιστική σημαίνει "πληροφορία" που αντλείται από απαριθμήσεις γεγονότων και χαρακτηριστικών ιδιοτήτων των ατόμων που μελετάμε.

Στο εξής τις "πληροφορίες" θα τις λέμε στατιστικά στοιχεία ή δεδομένα (data) ή εναλλακτικά παρατηρήσεις (observations).

Η Στατιστική (Statistics or Statistical Science) είναι η επιστήμη που πραγματεύεται, αναπτύσσει και εφαρμόζει επιστημονικές τεχνικές και μεθόδους συλλογής, οργάνωσης παρουσίασης και ανάλυσης δεδομένων. Στόχος της είναι αφενός η περιγραφή και η ερμηνεία καταστάσεων και φαινομένων και αφετέρου η συμπερασματολογία δηλαδή η εξαγωγή λογικών και τεκμηριωμένων συμπερασμάτων χρήσιμων στη λήψη ορθολογικών αποφάσεων.

§1.2 Χρησιμότητα της Στατιστικής

Δεν είναι υπερβολή να πούμε ότι η Στατιστική εφαρμόζεται σχεδόν παντού, σε όλους τους επιστημονικούς χώρους και τις ανθρώπινες δραστηριότητες (οικονομία, κοινωνικές επιστήμες,

Ιατρική, Επιδημιολογία, Βιολογία, Ψυχολογία, Οικολογία, Οικονομική
 ιατρική, σε όλους τους τομείς (δημόσιο, ιδιωτικό), αλλά και σε ατο-
 μικό επίπεδο. Δεν νοείται πια επιχείρηση που να μην χρη-
 σιμοποιεί στατιστικές μεθόδους για να ελέγξει την ποιότητα
 των προϊόντων που παράγει ή για να αναλύσει οικονομικούς
 παράγοντες ώστε να προγραμματίσει την παραγωγή της
 και το ανθρώπινο δυναμικό της. Η έρευνα αγοράς για τη
 μελέτη των συνθηκών των καταναλωτών έχει γίνει πλέον
 επιτακτική ενώ οι εταιρείες δημοσκοπήσεων (gallup) θα-
 ρβάνοντας ένα σχετικά μικρό αριθμό ψηφοφόρων μπορούν
 να προβλέπουν αποτελέσματα βουλευτικών και δημοτικών
 εκλογών.

§1.3 Σύνοψη Ιστορική Αναδρομή

Η Στατιστική με την έννοια της απλής συλλογής πληροφοριών
 απαραίτητων για τη λειτουργία του κράτους ήταν γνωστή από
 την αρχαιότητα. Η συσχέτιση των πληροφοριών γινόταν
 για στρατιωτικούς και φορολογικούς λόγους και είχε την
 έννοια της απογραφής (census, population census) της
 καθολικής καταγραφής του πληθυσμού, κυρίως ανδρών σε
 στρατεύσιμες ηλικίες. Για ορισμένους ερευνητές ο όρος
 "Στατιστική" προέρχεται από τη λατινική λέξη "status"
 η οποία εκτός από κατάσταση ή θέση αποδίδει και
 την έννοια της πολιτείας δηλαδή του κράτους. Σύμφωνα
 με ιστορικές μαρτυρίες, η πρώτη στοιχειώδης απογραφή
 διενεργήθηκε στην Κίνα το 2238 π. Χ. από τον αυτοκρά-
 τора Υαο. Κατόπιν πραγματοποιήθηκαν επίσης από

-4-

γραφές από τους Αιγύπτιους, τους Έλληνες και τους Ρωμαίους. Στο τέλος του Μεσαίωνα, στην Ευρώπη άρχισε η καταγραφή δηληριαρχιών στοιχείων (vital statistics) που αφορούν τους γάμους, τις γεννήσεις και τους θανάτους. Το ενδιαφέρον της πολιτείας, βαθμιαία επεντάθηκε στη συγμείνωση και άλλων πληροφοριών, κυρίως οικονομικού περιεχομένου. Κατά το 18^ο αιώνα άρχισε πιο συστηματικά η χρήση των στατιστικών μεθόδων δηλαδή η εφαρμογή ειδικών τεχνικών (statistical techniques) με σκοπό την εκτίμηση (estimation) και όχι απλά την καταγραφή και την απαρίθμηση (enumeration) των χαρακτηριστικών του πληθυσμού. Σταθροί στην ιστορία της Στατιστικής θεωρείται η εργασία του Graunt (1662) που αναφέρεται σε μία συστηματική ανάλυση των σωματικών θνησιμότητας ορισμένων περιοχών του Λονδίνου και του Halley (1693) ο οποίος μελετώντας την επιβίωση του πληθυσμού εισήγαγε στην ανάλυσή του την έννοια της τυχαιότητας (chance) και της πιθανότητας (probability). Η στοχαστική αυτή διάσταση και προσέγγιση των στατιστικών αναλύσεων αποτελεί τον θεμέλιο λίθο στον οποίο στηρίχθηκε το οικοδόμημα της Στατιστικής Θεωρίας. Η αλληγογραφία των Pascal (1623-1662) και Fermat (1601-1665) γύρω από τα τυχερά παιχνίδια εχαινίασε τον κλάδο των πιθανοτήτων.

Μέχρι τον 19^ο αιώνα πολύ σημαντική υπήρξε η συνεισφορά των Bayes (1702-1761), Gauss (1777-1855) και Laplace (1749-1827). Στη διάρκεια του 20^{ου} αιώνα θα κυριαρχήσουν τα ονόματα των Pearson, Markov, Neyman, Wald, Gini, Borel και Fisher.

§1.4 Πληθυσμός, Δείγμα και Στατιστική Συμπερασματολογία

Θεμελιώδη έννοια της Στατιστικής Θεωρίας είναι η έννοια του πληθυσμού (population). Ο όρος πληθυσμός χρησιμοποιείται για να δηλώσει οποιαδήποτε συλλογή ατόμων ζώων ή ρη. Μπορούμε π.χ. να θεωρήσουμε ανθρώπινους πληθυσμούς, πληθυσμούς φυτών, ζώων, λαδών κατά την αλυσή ενός κειμένου κλπ. Η Στατιστική ως επιστήμη που ασχολείται με δεδομένα που προκύπτουν από απλή απερίθμηση ή μέτρηση των ιδιοτήτων των πληθυσμών βασίζει την λειτουργία της στις στατιστικές μονάδες (statistical units) κάθε μία εκ των οποίων αποτελεί πηγή παραγωγής δεδομένων. Π.χ. αν θεωρήσουμε τον πληθυσμό των 12.000 φοιτητών του Πανεπιστημίου Πειραιώς, τότε κάθε φοιτητής αποτελεί και μία στατιστική μονάδα.

Η μελέτη του ύψους των φοιτητών του Πανεπιστημίου παράγει σύνολο 12.000 τιμών με συνέπεια ο πληθυσμός των φοιτητών να αντιστοιχεί σε πληθυσμό 12.000 μετρήσεων ύψους. Η μελέτη ενός τόσο μεγάλου αριθμού δεδομένων είναι συχνά δύσκολη στην πράξη.

Αυτή τη δυσκολία την ξεπερνάμε καταφεύγοντας στη μελέτη μέρους των τιμών του πληθυσμού. Το μέρος του πληθυσμού επιλέγεται με μεθόδους που υπάρχουν από τον κλάδο της δειγματοληψίας (sampling) και καλείται δείγμα (sample).

Μία ολοκληρωμένη στατιστική διαδικασία αποτελείται από τη λήψη δείγματος, την επεξεργασία των δεδομένων και την εξαγωγή συμπερασμάτων γύρω από τον πληθυσμό. Η επαγωγική διαδικασία της εξαγωγής συμπερασμάτων για τον πληθυσμό μέσα από την στατιστική επεξεργασία αποτελεί αντικείμενο της Επαγωγικής Στατιστικής ή Στατιστικής Συμπερασματολογίας. Η σχέση μεταξύ δείγματος και πληθυσμού είναι σχέση "μέρους προς όλο" και από τη μελέτη του μέρους (δείγματος) προκύπτουν συμπεράσματα περί του όλου (πληθυσμού). Για την περιγραφή του δείγματος χρησιμοποιούνται περιγραφικά μέτρα για τα οποία θα μιλήσουμε σε επόμενο κεφάλαιο. Τα αντίστοιχα αυτών των μέτρων στον πληθυσμό καλούνται παράμετροι (parameters). Στο παράδειγμα του ύψους των φοιτητών του Πανεπιστημίου Πειραιώς το μέσο ύψος στο δείγμα αποτελεί περιγραφικό στατιστικό μέτρο ενώ το άγνωστο μέσο ύψος στον πληθυσμό αποτελεί παράμετρο. Δηλαδή τα στατιστικά περιγραφικά μέτρα του δείγματος εστιμούν τις αντίστοιχες παραμέτρους του πληθυσμού.

Πληθυσμός με πεπερασμένο πλήθος στατιστικών μονάδων
καλείται πεπερασμένος (finite population). Σε κάθε
άλλη περίπτωση ο πληθυσμός είναι μη-πεπερασμένος
ή άπειρος (infinite population). Τυπικά, επειδή οι
ερευνητές συνήθως μελετούν πληθυσμούς μεγάλους σε
μέγεθος είναι σχεδόν αδύνατο να πάρουν τις επιθυμητές
πληροφορίες για κάθε ένα από τα μέλη τους. Καταφεύ-
γουν τις περισσότερες φορές στη δειγματοληψία.
Ο ερευνητής είναι αυτός που κάθε φορά ορίζει ποιος
είναι ακριβώς ο υπό μελέτη πληθυσμός ανάλογα
με τις ανάγκες και τις απαιτήσεις της μελέτης. Π.χ.
Θα ήταν δυνατόν να θεωρήσουμε ως υπό μελέτη πληθυσμό
"το σύνολο των φοιτητών του Πανεπιστημίου Πειραιώς"
ή "το σύνολο των πρωτοετών φοιτητών του Πανεπιστη-
μίου Πειραιώς" ή "το σύνολο των φοιτητών του Τμήματος
Πληροφορικής του Πανεπιστημίου Πειραιώς" κτλ. ανά-
λογα με το ποιο είναι το αντικείμενο μελέτης που
πραγματοποιούμε.

Αν επιλέξουμε το δείγμα μας ως εξής: τοποθετώντας
τα ονοματεπώνυμα των 12.000 εγγεγραμμένων
φοιτητών σε μία κάρτα υπό μορφή κλήρων, τους
οποίους ανακατεύουμε και επιλέγουμε με τυχαίο
τρόπο 600 κλήρους, το δείγμα το οποίο προκύπτει
μ' αυτό τον τρόπο καλείται απλό τυχαίο δείγμα
(random sample) και η διαδικασία επιλογής καλεί-
ται απλή τυχαία δειγματοληψία (random sampling).

Με την απλή τυχαία δειγματοληψία, όλες οι στατιστικές μονάδες που επιλέγουμε έχουν την ίδια δυνατότητα (πιθανότητα) επιλογής.

Ας θεωρήσουμε, τέλος, το εξής παράδειγμα. Εταιρεία δημοσκοπήσεων κάνει πανελλήνια πολιτική έρευνα και επιλέγει από το σύνολο των ψηφοφόρων δείγμα 1.000 ατόμων με σκοπό την πρόβλεψη του εκλογικού αποτελέσματος. Αν χρησιμοποιήσε την απλή τυχαία δειγματοληψία είναι δυνατόν π.χ. τα 850 άτομα που θα επιλεγούν από τα 1.000 να είναι κάτοικοι αγροτικών περιοχών. Σε μία τέτοια περίπτωση είναι αρτίστοχο αν οι προβλέψεις (επιτηρήσεις) που θα προκύψουν θα είναι σωστές. Συνεπώς για να καταλήξουμε σε σημαντικά και αξιόπιστα αποτελέσματα, το δείγμα πρέπει να είναι αντιπροσωπευτικό δηλαδή οι τυχές του δείγματος πρέπει να είναι τυπικές (αντιπροσωπευτικές) των τυχών του πληθυσμού ώστε τα αποτελέσματα της στατιστικής ανάλυσης να είναι αξιόπιστα.