

1. ΕΙΣΑΓΩΓΗ

Η γνώση μας, η συμπεριφορά μας, οι πράξεις και οι αποφάσεις μας, τόσο στην καθημερινή ζωή όσο και στην επιστημονική έρευνα, βασίζονται πάνω σε δείγματα. Πολύ συχνά, η πηγή των πληροφοριών για την διαμόρφωση γνώμης και την λήψη αποφάσεων είναι ένα πολύ μικρό δείγμα εμπειρίας (δύο ή τρεις επαφές με τους αρμόδιους μιας εταιρείας, δέκα μέρες σε μία ξένη χώρα κ.λ.π.). Άλλοτε, επειδή ο πληθυσμός (με την ευρεία έννοια του όρου) για τον οποίο απαιτούνται πληροφορίες είναι πεπερασμένος (οι κάτοικοι μιας πόλης), χρησιμοποιείται ολόκληρος ο πληθυσμός με βάση τα στοιχεία μιας απογραφής.

Ούτε ο πρώτος ούτε ο δεύτερος τρόπος αντιμετώπισης του προβλήματος είναι ενδεδειγμένος. Στην πρώτη περίπτωση, η έκταση της άγνοιας γύρω από τον πληθυσμό είναι τόσο μεγάλη, ώστε να καθιστά ανεπίτρεπτη οποιαδήποτε συμπερασματολογία στηρίζεται στο δείγμα. Στην δεύτερη περίπτωση, το κόστος μιας έρευνας από την άποψη χρόνου και χρήματος είναι μεγάλο. Μια "καλή" δειγματοληπτική μέθοδος φαίνεται να είναι η λύση που παρουσιάζει πολλά πλεονεκτήματα σε σύγκριση κυρίως με την απογραφή, όπως

α) **Χαμηλότερο κόστος:** Οι πληροφορίες στα δεδομένα προέρχονται από ένα τμήμα του πληθυσμού.

β) **Μεγαλύτερη ταχύτητα:** Η συλλογή και η επεξεργασία των δεδομένων είναι ταχύτερη. Αντίθετα, μια απογραφή είναι μια μη πρακτική λύση μια και τα συμπεράσματα μπορεί να μην είναι διαθέσιμα έγκαιρα. Άρα, ουσιαστικά, η εκλογή του ερευνητή είναι μεταξύ του να πάρει έγκαιρα την πληροφορία από ένα κατάλληλο δείγμα ή να μην κάνει τίποτα απολύτως.

γ) **Μεγαλύτερη ακρίβεια:** Πιο προσεκτική εποπτεία της διεξαγωγής της έρευνας και πιο προσεκτική επεξεργασία των αποτελεσμάτων είναι εφικτή.

Σήμερα, έχουν ξεπεραστεί και τα τελευταία ίχνη της καχυποψίας που υπήρχε ως προς την αξιοπιστία των πληροφοριών που περιέχονται σε ένα δείγμα και, έτσι, η δειγματοληψία είναι πια κοινή πρακτική σε πολλούς τομείς της καθημερινής ζωής. Κατασκευαστές και managers αποθηκών ενδιαφέρονται για την γνώμη και τις αντιδράσεις του καταναλωτικού κοινού για ένα προϊόν, καθώς και για τους λόγους για τους οποίους προτιμά ή όχι το προϊόν αυτό από κάποιο άλλο. Επίσης, επιχειρηματίες και βιομήχανοι χρησιμοποιούν δειγματοληπτικές μεθόδους για την αξιολόγηση της επάρκειας των εσωτερικών λειτουργιών των επιχειρήσεων ή βιομηχανιών, για την εκτίμηση των αποθεμάτων, την μελέτη της κατάστασης και της διάρκειας ζωής του εξοπλισμού τους, του ρυθμού παραγωγής εργασίας, για τον έλεγχο ποιότητας των προϊόντων και την μελέτη της κατανομής του χρόνου εργασίας των υπαλλήλων τους στα διάφορα καθήκοντά τους. Τέλος, έρευνες κοινής γνώμης πάνω σε τρέχοντα θέματα και έρευνες αγοράς γίνονται πολύ συχνά. Έτσι, η σημερινή κοινωνία κατακλύζεται από αριθμούς. Υπάρχει μια τάση να εκφράζονται ποσοτικά όλες οι πλευρές της καθημερινής μας ζωής, από τις διαιτητικές προτιμήσεις μέχρι τα πολιτικά ιδεώδη. Τα μέσα ενημέρωσης, περιλαμβανομένων και των διαφημίσεων, της τηλεόρασης, του ραδιοφώνου και των εφημερίδων, διατηρούν μια σταθερή ροή τέτοιων πληροφοριών. Καμία επιχειρηματολογία δεν είναι πλήρης χωρίς την υποστήριξη κάποιων αριθμών. Η παρουσίαση τέτοιων αριθμών έχει σκοπό να μας κρατήσει ενήμερους για την κατάσταση που επικρατεί γύρω μας και συχνά χρησιμοποιείται για την υποστήριξη κάποιας πρότασης ή και κριτικής, ή τουλάχιστον για την τοποθέτηση ενός θέματος στην σωστή του βάση. Τα αποτελέσματα της έρευνας αγοράς μπορούν, για παράδειγμα, να χρησιμοποιηθούν ως προβλέψεις του αποτελέσματος επικείμενων εκλογών ή για να τονίσουν την ανάγκη να μεταβληθεί η κείμενη

νομοθεσία σε σχέση με κάποιες διαμορφωθείσες κοινωνικές αντιλήψεις και απαιτήσεις.

Αναμφισβήτητα, τα άτομα είναι σήμερα περισσότερο από ποτέ άλλοτε πληροφορημένα, με την έννοια ότι δέχονται ποσοτικές περιγραφές του κόσμου μέσα στον οποίο ζουν. Το γεγονός αυτό είναι, βέβαια, πολύ καλό, αλλά επιβάλλει σοβαρές απαιτήσεις τόσο στον δέκτη όσο και στον πομπό των αριθμητικών πληροφοριών: από το ένα μέρος, ο απλός άνθρωπος χρειάζεται να έχει την ικανότητα να κατανοεί και να ερμηνεύει την πληροφόρηση, η οποία του παρέχεται. Από το άλλο μέρος, εκείνοι οι οποίοι παρουσιάζουν στατιστικά δεδομένα έχουν διπλή ευθύνη: να παρουσιάζουν τα στοιχεία με έντιμο και αντικειμενικό τρόπο και χωρίς την πρόθεση της παραπλάνησης του κοινού, ενώ, ταυτόχρονα, θα πρέπει να παρέχουν επαρκείς πληροφορίες σχετικά με την πηγή, τον στόχο και την μέθοδο της συλλογής των δεδομένων, τα οποία χρησιμοποίησαν, για την κατάλληλη ερμηνεία της οποιασδήποτε παραπέρα ανάλυσης.

Οι παραπάνω δύο απαιτήσεις δεν ικανοποιούνται πάντα. Παρά την εκπληκτική βελτίωση που έχει παρατηρηθεί τα τελευταία χρόνια στην ικανότητα κατανόησης θεμελιωδών ποσοτικών επιστημονικών εννοιών (numeracy), η στάση του απλού ανθρώπου απέναντι στα στατιστικά στοιχεία συχνά αντανακλά σύγχυση περισσότερο παρά κατανόηση ή διαφώτιση. Επιπλέον, η παρουσίαση των δεδομένων, όχι λιγότερο συχνά, στερείται ακρίβειας και σαφήνειας. Ανακριβείς ή ελλιπείς διατυπώσεις, άκυρη συμπερασματολογία, γραφήματα με παραποιημένες ή μη προσδιοριζόμενες κλίμακες (σχετιζόμενα με μη καλώς ορισμένες έννοιες) και εικονογράμματα, με ψυχολογικό αντίκτυπο διαφορετικό από αυτό της πραγματικής τους βάσης, προκαλούν σύγχυση στον δέκτη των πληροφοριών. Τέτοιας μορφής πληροφόρηση, ανεξάρτητα από το εάν γίνεται ηθελημένα ή είναι αποτέλεσμα έλλειψης στατιστικής εμπειρογνωμοσύνης, εξυπηρετεί την προώθηση διαφόρων οικονομικών συμφερόντων. Βέβαια, τα πρότυπα (standards) βελτιώνονται συνεχώς και οι χειρότερες υπερβάσεις τους ανήκουν ίσως στο πιο μακρινό παρελθόν.

Τα παραπάνω τονίζουν την ανάγκη μεγαλύτερης προσοχής στην ερμηνεία στατιστικών πληροφοριών, με βάση την ορθή γνώση και την εφαρμογή καταλλήλων μεθόδων συλλογής, παρουσίασης και ερμηνείας στατιστικών δεδομένων.

Ο στόχος αυτού του βιβλίου είναι η παροχή τέτοιας γνώσης και ικανότητας. Η παρουσίαση προϋποθέτει ότι ο αναγνώστης έχει στοιχειώδη γνώση εννοιών και μεθόδων των Πιθανοτήτων και της Στατιστικής.

1.1 Πεπερασμένοι Πληθυσμοί, Δειγματοληπτικές Έρευνες και Σφυγμομετρήσεις (Finite Populations, Sample Surveys and Opinion Polls)

Το κοινό χαρακτηριστικό των περιπτώσεων που αναφέρθηκαν στα προηγούμενα είναι ότι όλες συνδέονται με *πεπερασμένους πληθυσμούς*, οι οποίοι περιέχουν ένα περιορισμένο και καλώς ορισμένο σύνολο *ατόμων* ή *ομάδων*: οι τιμές ενοικιαζόμενων διαμερισμάτων στην ευρύτερη περιοχή της Αθήνας για τους τελευταίους τέσσερις μήνες, οι άνθρωποι ανά τον κόσμο που έχουν πρόσβαση σε πληροφόρηση μέσω της τηλεόρασης, το σύνολο των εργαζομένων στην περιοχή της Αττικής, το σύνολο των σπύρων σε ένα σπυρτόκουτο κ.ο.κ.. Ο πληθυσμός μπορεί να είναι πολύ μικρός σε μέγεθος (όπως, για παράδειγμα, τα σπύρα σε ένα σπυρτόκουτο) ή πολύ μεγάλος (αρκετά εκατομμύρια), αλλά είναι πεπερασμένος.

Ανεξάρτητα από το μέγεθος του πληθυσμού, ο στόχος είναι η συμπερασματολογία σχετικά με αυτόν τον πεπερασμένο πληθυσμό με βάση τη συλλογή και ανάλυση πληροφοριών που σχετίζονται (κατά κύριο λόγο) με ένα μέρος του – αυτό που ονομάζουμε **δείγμα** από τον πληθυσμό. Οι πληροφορίες αυτές συλλέγονται από έρευνα και επισκόπηση (survey) του πληθυσμού. Η μελέτη του τρόπου διεξαγωγής τέτοιων **δειγματοληπτικών ερευνών (sample surveys)** αποτελεί και το θέμα του παρόντος βιβλίου.

Οι δειγματοληπτικές έρευνες μπορούν να καλύψουν οποιοδήποτε θέμα συνδέεται με τα χαρακτηριστικά ενός πεπερασμένου πληθυσμού, αλλά μία ευρεία κατηγοριοποίηση των περιοχών των διαφορετικών αυτών θεμάτων θα ήταν χρήσιμη. Οι περισσότερες δειγματοληπτικές έρευνες έχουν στόχο την περιγραφή ανθρώπινων πληθυσμών και του περιβάλλοντός τους και διακρίνονται στις εξής κατηγορίες:

- ο Δημογραφικά χαρακτηριστικά του πληθυσμού (π.χ. μέγεθος οικογενειών)
- ο Οικονομική δομή της κοινωνίας (π.χ. βιομηχανική δραστηριότητα)
- ο Τρόπος ζωής (π.χ. ψυχαγωγία)
- ο Κοινωνικό περιβάλλον (π.χ. ύψος εισοδήματος)
- ο Απόψεις και γνώμες (π.χ. στάση απέναντι στην κυβερνητική πολιτική).

Στο σημείο αυτό θα ήταν χρήσιμες μερικές βασικές παρατηρήσεις. Κατά κανόνα, η έρευνα θα μπορούσε να βασιστεί στην επισκόπηση όλων των μελών ενός πεπερασμένου πληθυσμού. Μια τέτοια διαδικασία ονομάζεται **απογραφή (census)**. Η απογραφή είναι, δηλαδή, μια δειγματοληπτική έρευνα με κάλυψη 100%. Το ενδιαφέρον μας, όμως, εστιάζεται σε πολύ χαμηλότερα επίπεδα κάλυψης, συχνά του ύψους του 1% ή 5%.

Συχνά, οι πληροφορίες που συλλέγονται πάνω στα άτομα ενός πληθυσμού είναι ποσοτικές και περιγράφουν κοινωνικά ή οικονομικά χαρακτηριστικά. Σε άλλες περιπτώσεις, ενδέχεται να είναι ποιοτικές ή ακόμη και υποκειμενικές, με την μορφή προσωπικών απόψεων ή προτιμήσεων. Στην τελευταία περίπτωση, η δειγματοληπτική έρευνα είναι γνωστή ως **σφυγμομέτρηση της κοινής γνώμης (opinion poll)**.

Αν το ενδιαφέρον μας εστιάζεται στην ποιότητα προϊόντων ή στην συμπεριφορά καταναλωτών, η δειγματοληπτική έρευνα είναι γνωστή ως **έρευνα αγοράς (market research)**.

Οι κανόνες και οι μέθοδοι συλλογής και ανάλυσης δεδομένων από πεπερασμένους πληθυσμούς αποτελεί την περιοχή της Στατιστικής που είναι γνωστή ως **Μέθοδοι Δειγματοληπτικών Ερευνών (Sample**

Survey Methods). Η θεωρητική τους βάση ονομάζεται **Θεωρία Δειγματοληψίας (Sampling Theory)**. Οι δειγματοληπτικές έρευνες μπορούν παραπέρα να διακριθούν σε δύο κατηγορίες: **Περιγραφικές** και **Αναλυτικές**. Στην πρώτη κατηγορία, στόχος είναι η άντληση πληροφοριών για μεγάλες ομάδες (π.χ. για τον αριθμό ανδρών, γυναικών ή παιδιών που παρακολουθούν ένα πρόγραμμα τηλεόρασης). Στην δεύτερη κατηγορία, σκοπός είναι η συναγωγή συμπερασμάτων με βάση τις πληροφορίες από το δείγμα και η συγκριτική μελέτη αποτελεσμάτων που προέρχονται από δύο ή περισσότερες υποομάδες του πληθυσμού για την ανίχνευση τυχόν διαφορών και τον έλεγχο υποθέσεων σχετικά με τους λόγους ύπαρξης των διαφορών αυτών.

Στα κεφάλαια που ακολουθούν, θα αναπτυχθούν διάφορες δειγματοληπτικές τεχνικές και το απαραίτητο θεωρητικό υπόβαθρο για την συναγωγή συμπερασμάτων με βάση τις πληροφορίες του εκάστοτε τυχαίου δείγματος.

Οι δειγματοληπτικές αυτές τεχνικές έχουν ευρεία πρακτική εφαρμογή σε περιοχές όπως η Γεωργία, η Εκπαίδευση, η Βιομηχανία, οι Κοινωνικές Επιστήμες, η Ιατρική και οι Υπηρεσίες Κοινή Ωφέλειας.

Η επιλογή της δειγματοληπτικής τεχνικής που θα χρησιμοποιηθεί, όπως και του μεγέθους του δείγματος που απαιτείται, επηρεάζεται από ορισμένα πρακτικά προβλήματα (όχι απαραίτητα στατιστικής φύσης), τα οποία αναφέρονται, όταν από τον σχεδιασμό της δειγματοληπτικής έρευνας μετακινηθούμε στην διεξαγωγή της. Τέτοια προβλήματα αναφέρονται σε πρακτικές πλευρές της έρευνας, όπως:

- ο Ο ορισμός του πληθυσμού και του χαρακτηριστικού που επιδιώκεται να εκτιμηθεί
- ο Η μέθοδος της λήψης του δείγματος (δημοσιευμένα στοιχεία, συνεντεύξεις, ταχυδρομικές ή τηλεφωνικές συνεντεύξεις)
- ο Η μεροληπτικότητα των απαντήσεων ή των μη-απαντήσεων (response and non-response bias)
- ο Οι πιλοτικές έρευνες (pilot surveys)
- ο Ο σχεδιασμός του ερωτηματολογίου και η διατύπωση των ερωτήσεων

- ο Η τεχνική της συνέντευξης
- ο Ο προϋπολογισμός της έρευνας
- ο Η χρήση συμπληρωματικής πληροφόρησης
- ο Σύνθετες δειγματοληπτικές έρευνες

Μερικά από αυτά τα πρακτικά θέματα εξετάζονται στο τελευταίο κεφάλαιο του παρόντος βιβλίου (κεφάλαιο 7), ενώ άλλα θίγονται σε κατάλληλα στάδια της παρουσίασης των διαφόρων δειγματοληπτικών τεχνικών.

1.2 Ιστορική Αναδρομή

Η τεχνική της συλλογής πληροφοριών με βάση την άμεση επαφή με τα άτομα ενός πληθυσμού έχει μακρά ιστορία. Οι αρχαίες αυτοκρατορίες της Αιγύπτου και της Ρώμης χρησιμοποιούσαν περιοδικές απογραφές ως βάση για τον καθορισμό των φόρων, την στρατολόγηση και άλλες διοικητικές αποφάσεις. Ήταν, όμως, αρχές του 18ου αιώνα, όταν για πρώτη φορά εφαρμόστηκε έρευνα ευρείας κλίμακας με οργανωμένο τρόπο για να μελετηθούν κοινωνικά προβλήματα. Ο Βρετανός μεταρρυθμιστής John Howard ήταν ένας από τους πρωτοπόρους σε αυτή την προσπάθεια με την λεπτομερή έρευνα που έκανε για την επίδραση των συνθηκών των φυλακών στην υγεία των φυλακισμένων. Η πρώτη του αυτή μελέτη άρχισε την δεκαετία του 1770 στην Αγγλία και επεκτάθηκε αργότερα σε άλλες χώρες σε μια προσπάθεια που θα μπορούσε να χαρακτηριστεί ως η πρώτη διεθνική εφαρμογή των δειγματοληπτικών μεθόδων. Ο Frederic LePlay, ένας Γάλλος οικονομολόγος του 19ου αιώνα, κατέβαλε μεγάλη προσπάθεια να χρησιμοποιήσει δειγματοληπτικές έρευνες για τον ορθολογικό κοινωνικό σχεδιασμό. Η μελέτη του για το εισόδημα και την κατανάλωση ευρωπαϊκών οικογενειών αντανάκλα μια θαυμαστή ισορροπία μεταξύ μεταρρυθμιστικού ζήλου και επιστημονικής αντικειμενικότητας. Οι προσπάθειές του να ελέγξει τις απαντήσεις στο ερωτηματολόγιο έναντι της πληροφόρησης που είχε από ανεξάρτητες

πηγές, όπως η άμεση παρατήρηση και οι αναφορές άλλων ερωτηθέντων, αντανακλούν μια μάλλον επιτηδευμένη γνώση της ερευνητικής μεθοδολογίας. Βέβαια, οι εκτενείς και με πολλούς στόχους δειγματοληπτικές έρευνες του σήμερα έχουν τις ρίζες τους στις μελέτες του Άγγλου στατιστικού Charles Booth. Το 1866, ο στατιστικός αυτός ανέλαβε μια μεγάλη μελέτη της φτώχειας, η οποία τελικά δημοσιεύθηκε σε 17 τόμους το 1897.

Από γεωγραφική άποψη, η πρωτοπορία σε παγκόσμια επίπεδα όσον αφορά την θεωρία και τις εφαρμογές της δειγματοληψίας που είναι βασισμένη στις πιθανότητες, έχει περάσει από τρεις περιόδους:

1. Από την αρχή του 19ου αιώνα, όταν ο Gauss, ο Laplace, ο Legendre και οι άλλοι εισήγαγαν την μαθηματική θεωρία του "συνδυασμού των παρατηρήσεων" (combination of observations), μέχρι περίπου το 1890, η πρωτοπορία βρισκόταν στην ηπειρωτική Ευρώπη.

2. Από το 1890, περίπου, όταν ο Karl Pearson (ο οποίος πείσθηκε από τον Francis Galton), ο G. Udny Yule και οι άλλοι επιστήμονες ανέπτυξαν την μεθοδολογία της παλινδρόμησης και της συσχέτισης και, αργότερα, όταν ο R.A. Fisher, ο Jersey Neyman και άλλοι εισήγαγαν την θεωρία των μικρών δειγμάτων και την δειγματοληψία από πεπερασμένους πληθυσμούς, μέχρι περίπου την αρχή του β' παγκοσμίου πολέμου, η επιστημονική πρωτοπορία ανήκε στην Μεγάλη Βρετανία.

3. Από το 1940, περίπου, η πρωτοπορία αυτή πέρασε στις Ηνωμένες Πολιτείες. Αυτό συνέβη όταν στατιστικοί που υπηρετούσαν στην κυβέρνηση των Ηνωμένων Πολιτειών, με κορυφαίους τους Morris H. Hansen και William N. Hurwitz ανέπτυξαν μεθόδους δειγματοληπτικών ερευνών και θεωρίες για ανθρώπινους πληθυσμούς. Η περίοδος αυτή επηρεάστηκε επίσης σημαντικά από πανεπιστημιακούς, με κορυφαίους τον Harold Hotelling, τον George W. Snedecor, τον Abraham Wald, τον S. S. Wilks και άλλους, οι οποίοι συνέβαλαν ουσιαστικά στην πρόοδο της θεωρίας της δειγματοληψίας.

Σε κάθε μια από τις τρεις αυτές περιόδους, οι στατιστικοί αντιμετώπισαν και προσπάθησαν να επιλύσουν διαφορετικά προβλήματα

και διαμόρφωσαν διαφορετικές θεωρητικές δομές. Γενικά, όμως, μπορεί να λεχθεί ότι κατά τη διάρκεια της 2ης και της 3ης περιόδου, οι θεωρητικοί επιστήμονες διαμόρφωσαν συγκροτημένες θεωρίες βασισμένες στα αποτελέσματα προηγούμενων εργασιών. Πράγματι, ακόμα και την πρώτη περίοδο, τα αποτελέσματα βασίσθηκαν στους νόμους των πιθανοτήτων που είχαν ήδη εισαχθεί από τον Bernoulli και άλλους. Η συνεισφορά του Gauss και άλλων επιστημόνων στην αρχή του 19ου αιώνα αναφέρεται στην εκτίμηση των τροχιών των πλανητών στον ουρανό από αστρονομικές παρατηρήσεις. Παρ' όλα αυτά, οι μέθοδοι που αναπτύχθηκαν από τον Gauss βρήκαν σύντομα εφαρμογή σε ένα ευρύ φάσμα επιστημονικών δεδομένων που περιλάμβαναν ακόμη και οικονομικά και κοινωνικά δεδομένα. Η αρχική συνεισφορά του Karl Pearson και των συνεργατών του, όπως επίσης και ένα σημαντικό μέρος της εργασίας που έκαναν αργότερα αναφερόταν σε προβλήματα της Γενετικής, της Ανθρωπολογίας και άλλων τομέων των Ανθρωπιστικών Επιστημών. Η εργασία των στατιστικών που εργάστηκαν για την αμερικανική κυβέρνηση στην πιο πρόσφατη περίοδο, είχε ως στόχο τον προσδιορισμό αξιόπιστων εκτιμήσεων των οικονομικών και δημογραφικών παραμέτρων για χρήση, από την αμερικανική κυβέρνηση κατά την διάρκεια της μεγάλης οικονομικής ύφεσης του β' παγκοσμίου πολέμου και αργότερα.

Δεν αποτελεί, επομένως, έκπληξη ότι προβλήματα τόσο διαφορετικής φύσης απαιτούν διαφορετικές θεωρητικές συλλήψεις και μαθηματικές αντιμετώψεις. Εκείνο που αποτελεί έκπληξη είναι ότι τόσο διαφορετικές μαθηματικές προσεγγίσεις μπορούν να περιληφθούν από κοινού σ' αυτό που σήμερα είναι γνωστό ως θεωρία και πρακτική της δειγματοληψίας που στηρίζεται στην Θεωρία Πιθανοτήτων (probability sampling).

Θα πρέπει, ίσως, να επισημανθεί ότι κατά τις τρεις διαφορετικές περιόδους, στις οποίες αναφερθήκαμε προηγουμένως, ο δυτικός πολιτισμός βρέθηκε σε τρεις διαφορετικές φάσεις ανάπτυξης, γεγονός που είχε σημαντική επίδραση στην ανάπτυξη της Στατιστικής.

1. Η περίοδος των Gauss-Laplace αποτελούσε μέρος μιας έκρηξης πνευματικής άνθισης που συνόδευσε την Γαλλική Επανάσταση και την περίοδο του Ναπολέοντα. Η περίοδος αυτή των Gauss-Laplace χαρακτηρίζεται από μια σειρά επιτευγμάτων της εποχής στις φυσικές επιστήμες.

2. Η περίοδος των Pearson-Fisher συνόδευσε μια εντατική βιομηχανοποίηση (που ήταν, ίσως, η πιο σημαντική εφαρμογή της προόδου στις φυσικές επιστήμες) στην Ευρώπη και εξαπλώθηκε στον υπόλοιπο κόσμο. Η σχέση της, όμως, με την εξέλιξη αυτή δεν ήταν πάρα πολύ στενή. Πολύ πιο σημαντική και προφανής ήταν η σχέση της με την εντυπωσιακή ανάπτυξη που συνέβη την ίδια περίοδο στις Ανθρωπιστικές Επιστήμες και στην Γεωργία.

3. Η περίοδος των Hansen-Hurwitz ήταν αναπόσπαστο μέρος της πολιτικής της κυβέρνησης των Ηνωμένων Πολιτειών α) της ανασυγκρότησης από τη μεγάλη ύφεση, β) της διεξαγωγής και της νίκης στον β' παγκόσμιο πόλεμο και γ) της διαχείρισης και της ανάπτυξης της μεταπολεμικής περιόδου.

1.3 Βασικές Έννοιες και Ορισμοί

Στην ενότητα αυτή, δίνονται ορισμοί βασικών εννοιών, στις οποίες γίνεται εκτενής αναφορά στα κεφάλαια που ακολουθούν.

Θεμελιώδης έννοια στην παρουσίαση του προβλήματος της δειγματοληψίας είναι η έννοια του **πεπερασμένου πληθυσμού**. Τα **άτομα** του πληθυσμού (ο οποίος δεν είναι απαραίτητα ανθρώπινος) έχουν ορισμένα μέτρα, των οποίων η εκτίμηση παρουσιάζει ενδιαφέρον. Συχνά, ενδιαφερόμαστε να μελετήσουμε ορισμένα **χαρακτηριστικά** του πληθυσμού ως προς κάποιο μέτρο, όπως ο **μέσος** της διάρκειας ζωής ενός τύπου ηλεκτρικού λαμπτήρα, το **συνολικό ύψος** της παραγωγής λαδιού σε μια περιοχή ή το **ποσοστό** των οικογενειών με εισοδήματα που υπερβαίνουν τα 18 χιλιάδες ευρώ κατά την διάρκεια του 2002.

Κατά την διεξαγωγή μιας δειγματοληπτικής έρευνας, συχνά, δημιουργούνται ασάφειες όσον αφορά την ερμηνεία ορισμένων βασικών εννοιών. Πιο συγκεκριμένα, υπάρχει κάποια ασυνέπεια στον τρόπο με τον οποίο χρησιμοποιούνται οι διαφορετικές έννοιες. Για τον λόγο αυτό, παραθέτουμε στην συνέχεια μερικές από τις συνηθέστερες βασικές έννοιες.

Αντικειμενικός Πληθυσμός (Target Population): Είναι το σύνολο του πεπερασμένου πληθυσμού για τον οποίο ενδιαφερόμαστε να συλλέξουμε πληροφορίες. Για παράδειγμα, ο αντικειμενικός πληθυσμός μιας μελέτης, θα μπορούσε να είναι το σύνολο των εργαζομένων στην Ελλάδα που είναι κάτω από 30 ετών.

Υπό Μελέτη Πληθυσμός (Study Population): Είναι το βασικό πεπερασμένο σύνολο ατόμων που θέλουμε να μελετήσουμε. Για παράδειγμα, το σύνολο των εργαζομένων στην Ελλάδα ηλικίας κάτω των 30 ετών, των οποίων η μόνιμη κατοικία είναι μέσα στην περιοχή του λεκανοπεδίου Αττικής. Ο πληθυσμός αυτός ενδέχεται να ταυτίζεται με τον αντικειμενικό πληθυσμό. Συχνά, όμως, ενδέχεται να είναι ένας πληθυσμός περισσότερο περιορισμένος σε μέγεθος, αλλά και περισσότερο εύκολο να μελετηθεί, του οποίου οι ιδιότητες ελπίζεται ότι μπορούν να προβληθούν στον ευρύτερο αντικειμενικό πληθυσμό.

Χαρακτηριστικό του Πληθυσμού (Population Characteristic): Είναι η πλευρά του πληθυσμού που επιθυμούμε να μελετήσουμε (να εκτιμήσουμε). Για παράδειγμα, το ποσοστό των πτήσεων ενός διεθνούς αεροδρομίου που εκτελούνται χωρίς καθυστέρηση, ή το σύνολο της παραγωγής κρασιού στην Ελλάδα το 2000, ή το μέσο ημερομίσθιο ενός ανειδίκευτου εργάτη κλωστοϋφαντουργίας.

Δειγματοληπτικές Μονάδες (Sampling Units): Είναι οι μονάδες που επιλέγουμε με σκοπό να επιτύχουμε πρόσβαση στα άτομα (μονάδες) του υπό μελέτη πληθυσμού. Για παράδειγμα, ας υποθέσουμε ότι επιθυμούμε να κάνουμε μια έρευνα οικογενειακών προϋπολογισμών σε κάποια πόλη. Αν και είναι προφανές ότι τα άτομα (οι μονάδες) του υπό μελέτη πληθυσμού είναι *οικογένειες*, πρέπει να υιοθετηθεί κάποιος συμβατικός ορισμός της έννοιας της οικογένειας πριν προχωρήσουμε

στην έρευνα. Ακόμα και σε αυτή την περίπτωση ενδέχεται να μην υπάρχει εύκολος τρόπος δειγματοληψίας τέτοιων μονάδων (οικογενειών).

Θα ήταν, ίσως, ευκολότερο να ληφθεί ένα δείγμα διευθύνσεων και να αναζητηθούν πληροφορίες για τις οικογένειες που κατοικούν στις επιλεγείσες διευθύνσεις. Έτσι, οι διευθύνσεις αποτελούν τις δειγματοληπτικές μονάδες, παρά το γεγονός ότι ο πληθυσμός των διευθύνσεων δεν είναι ουσιώδους σημασίας.

Δειγματοληπτικό Πλαίσιο (Sampling Frame): Είναι το σύνολο των δειγματοληπτικών μονάδων, το οποίο αποτελεί και την πηγή του δείματός μας. Μερικές φορές, οι δειγματοληπτικές μονάδες ταυτίζονται με τα μέλη του υπό μελέτη πληθυσμού. Στις περισσότερες περιπτώσεις, όμως, αυτό δεν συμβαίνει και έτσι το δειγματοληπτικό πλαίσιο είναι μία λιγότερο ακριβής υποδιαίρεση του υπό μελέτη πληθυσμού (coarser subdivision), κάθε μονάδα της οποίας (δειγματοληπτική μονάδα) περιέχει ένα διακεκριμένο σύνολο μελών του πληθυσμού.

Κατάλογος ή Λίστα (List): Για να χρησιμοποιηθεί το δειγματοληπτικό πλαίσιο ως η πρωταρχική πηγή από την οποία θα επιλεγεί το δείγμα, θα πρέπει να είναι δυνατός ο προσδιορισμός των δειγματοληπτικών μονάδων. Εξάλλου, το δειγματοληπτικό πλαίσιο επιλέγεται με αυτό τον σκοπό. Στην καλύτερη περίπτωση, ένας πραγματικός κατάλογος όλων των δειγματοληπτικών μονάδων ενδέχεται να υπάρχει, όπως, για παράδειγμα, ο κατάλογος των διευθύνσεων μιας πόλης ή ο κατάλογος που παρέχεται από ιατρικά αρχεία για όλους τους ηλικιωμένους ασθενείς, οι οποίοι επισκέπτονται τους γιατρούς τους σε μια συγκεκριμένη περιοχή και για μια ορισμένη χρονική περίοδο. Ένας τέτοιος κατάλογος διευκολύνει ιδιαίτερα την επιλογή του δείματος.

1.4 Περιπτώσεις Λήψης Δείματος

Το φυσικό ερώτημα που ανακύπτει είναι πώς επιλέγεται ένα δείγμα. Με την απάντηση στο ερώτημα αυτό ασχολούνται τα κεφάλαια

που ακολουθούν. Στην παρούσα ενότητα, θα μπορούσαμε, ίσως, να αναφερθούμε στο πρόβλημα αυτό σε μια καθαρά διαισθητική βάση.

Ο κύριος στόχος είναι η λήψη ενός δείγματος, το οποίο να είναι αντιπροσωπευτικό του πληθυσμού και το οποίο να οδηγεί σε εκτιμήσεις των χαρακτηριστικών του πληθυσμού με όσο το δυνατόν μεγαλύτερη ακρίβεια μπορούμε να επιτύχουμε για το κόστος ή για την προσπάθεια που είμαστε έτοιμοι να καταβάλουμε.

Μερικά δειγματοληπτικά σχήματα που έχουν διαισθητική βάση και που αναπτύχθηκαν για να αντιμετωπίσουν συγκεκριμένα προβλήματα (ad hoc schemes), εξελίχθηκαν σε δημοφιλή σχήματα και χρησιμοποιούνται ευρέως. Τέτοια δειγματοληπτικά σχήματα περιλαμβάνουν τα εξής:

Δειγματοληψία Ευκολίας (Προσιτότητας) ή Πρόχειρη Δειγματοληψία (Accessibility ή Haphazard Sampling): Το δείγμα επιλέγεται από ένα τμήμα του πληθυσμού, στο οποίο υπάρχει εύκολη πρόσβαση. Δηλαδή, επιλέγονται μόνο εκείνες οι μονάδες του πληθυσμού που είναι ευκολότερο να ληφθούν. Είναι φανερό ότι ένα τέτοιο δείγμα δεν μπορεί να είναι αντιπροσωπευτικό του πληθυσμού από τον οποίο προέρχεται. Ενδεικτικά, αναφέρουμε μερικές περιπτώσεις, οι οποίες οδηγούν σε δείγματα τέτοιας μορφής:

(α) Για τον έλεγχο των πορτοκαλιών που παράγει ένας παραγωγός επιλέγονται μερικά πορτοκάλια από ένα χαρτοκιβώτιο από τα σημεία του χαρτοκιβωτίου που μπορεί να φθάσει το χέρι.

(β) Σε μια σφυγμομέτρηση της κοινής γνώμης, αναζητούνται απαντήσεις εθελοντών στο ερωτηματολόγιο.

(γ) Οι αναγνώστες ενός περιοδικού καλούνται να συμπληρώσουν να να επιστρέψουν ένα ερωτηματολόγιο που επισυνάπτεται στο περιοδικό.

(δ) Το ερωτηματολόγιο που εστάλη από το Υπουργείο Παιδείας στο πλαίσιο του εθνικού διαλόγου για την παιδεία το 1991.

Τα μειονεκτήματα τέτοιων δειγμάτων, όσο αφορά τη χρησιμοποίησή τους για την συναγωγή συμπερασμάτων για ολόκληρο

τον πληθυσμό, είναι προφανή. Δυστυχώς, σε άλλες περιπτώσεις ενδέχεται να μην είναι προφανή αν και όχι λιγότερο σοβαρά!

Δειγματοληψία Κρίσης ή Σκοπιμότητας (Judgemental or Purposive Sampling): Στις περιπτώσεις που ο πληθυσμός ενδέχεται να αποτελείται από διαφορετικούς τύπους μονάδων, με διαφορετικούς βαθμούς προσιότητας, ο ερευνητής προβαίνει σε *εσκεμμένη υποκειμενική επιλογή* κατά την λήψη αυτού που θεωρεί *αντιπροσωπευτικό δείγμα*. Ο ερευνητής, δηλαδή, επιλέγει από τον ανομοιογενή πληθυσμό ένα δείγμα *τυπικών μονάδων*, δηλαδή μονάδων οι οποίες βρίσκονται κοντά στην κατά την αντίληψή του μέση τιμή του πληθυσμού. Τα αποτελέσματα μιας τέτοιας μορφής δειγματοληψίας μπορεί να είναι πολύ καλά, αν η διαίσθηση του ερευνητή ή η κρίση του είναι τέλεια. Θα πρέπει βέβαια να σημειωθεί ότι αρκετές δειγματοληπτικές έρευνες υιοθετούν αυτό το σχήμα.

Δειγματοληψία με Προκαθορισμένα Ποσοστά (Quota Sampling): Το σχήμα αυτό είναι αποτέλεσμα του συνδυασμού κρίσης του ερευνητή και προσιότητας των μονάδων του πληθυσμού. Για παράδειγμα, στην δειγματοληψία με προκαθορισμένα ποσοστά (βλέπε κεφάλαιο 4), ενδέχεται να ερωτηθούν οι περαστικοί ενός δρόμου σε μια προσπάθεια να ληφθεί ένα δείγμα που κρίνεται από τον ερευνητή ότι είναι αντιπροσωπευτικό διαφορετικών ηλικιών, φύλου, επαγγέλματος κ.ο.κ. Το σχήμα, όμως, αυτό εμπλέκει ένα στοιχείο προσιότητας: *Επιλέγονται "οι περαστικοί που κρίνονται ότι ανταποκρίνονται περισσότερο από τους υπόλοιπους στα παραπάνω κριτήρια"*.

Ενώ οι μέθοδοι αυτές ενδέχεται να οδηγούν σε αντιπροσωπευτικά δείγματα, τα οποία κάτω από τις κατάλληλες συνθήκες μπορεί να δώσουν χρήσιμα αποτελέσματα, δεν επιδέχονται ανάπτυξη δειγματοληπτικής θεωρίας. Ο λόγος είναι ότι στερούνται του στοιχείου της τυχαιότητας κατά την επιλογή των μονάδων. Ως αποτέλεσμα, δεν υπάρχει μέτρο σύγκρισης της αντιπροσωπευτικότητας ή της καταλληλότητας ή της ακρίβειας των εκτιμητριών που θα βασισθούν σε δείγματα που έχουν προκύψει από τα παραπάνω σχήματα. Επομένως, επιβάλλεται να εισαχθεί ένα στοιχείο τυχαιότητας στην διαδικασία

επιλογής δείγματος, ώστε το δείγμα να επιλέγεται σύμφωνα με κάποιο μηχανισμό πιθανότητας (probability mechanism). Μια τέτοια αντιμετώπιση δίνει την δυνατότητα να ελεγχθεί η αντιπροσωπευτικότητα και να εκτιμηθεί η ακρίβεια της δειγματοληπτικής έρευνας και των προκυπτουσών εκτιμητριών. Η εισαγωγή του μηχανισμού πιθανότητας μπορεί να επιτευχθεί ως εξής:

Έστω n το μέγεθος του δείγματος που επιθυμούμε να επιλέξουμε. Θεωρούμε, στην συνέχεια, (εννοιολογικά τουλάχιστον) όλα τα δυνατά δείγματα μεγέθους n που θα μπορούσαν να επιλεγούν από τον συγκεκριμένο πληθυσμό, έστω S_1, S_2, \dots . Δηλαδή, κάθε S_i συμβολίζει ένα διακεκριμένο δείγμα μεγέθους n από ολόκληρο τον πληθυσμό.

Η Δειγματοληψία κατά Πιθανότητα (Probability Sampling) βασίζεται στην αντιστοίχιση μιας πιθανότητας π_i στο δείγμα S_i , $i = 1, 2, \dots$ και στην εν συνεχεία επιλογή ενός δείγματος, S , από το σύνολο των δυνατών δειγμάτων S_1, S_2, \dots σύμφωνα με ένα μηχανισμό πιθανότητας, ο οποίος εξασφαλίζει στο δείγμα S_i πιθανότητα επιλογής ίση με π_i ($i=1,2,\dots$). Είναι προφανές, ότι υπάρχει μια απειρία διαφορετικών κατά πιθανότητα δειγματοληπτικών σχημάτων αντιστοιχούντων στις διαφορετικές κατανομές πιθανότητας $\pi = \{\pi_1, \pi_2, \dots\}$ που μπορούν να ορισθούν πάνω στο σύνολο των δυνατών δειγμάτων S_1, S_2, \dots . Τα σχήματα αυτά μπορούν να είναι πάρα πολύ απλά έως πάρα πολύ επιτηδευμένα και σύνθετα.

Στα κεφάλαια που ακολουθούν θεωρούμε μερικά από τα ευρύτερα χρησιμοποιούμενα σχήματα, συγκρίνοντάς τα ως προς την ακρίβεια της εκτίμησης των παραμέτρων που ενδιαφερόμαστε να μελετήσουμε.