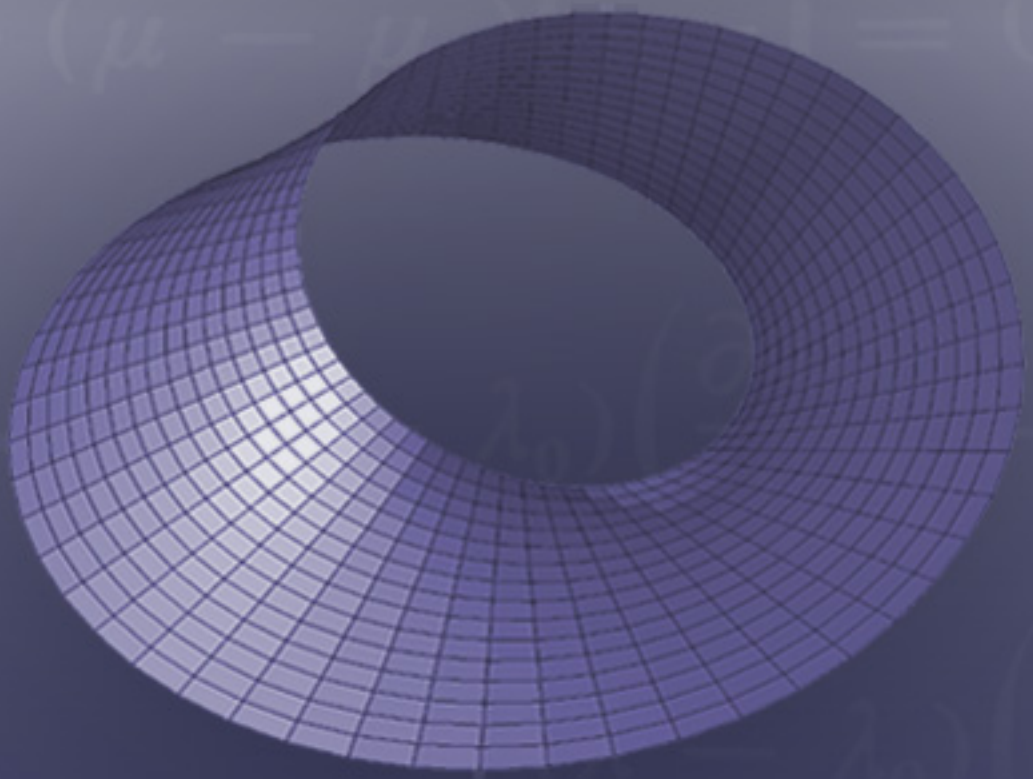
Πληθυσμός και χαρακτηριστικά



ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ

ΠΛΗΘΥΣΜΟΣ ΚΑΙ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

1.1. *Πληθυσμός και Στατιστικές Μονάδες*

Τα σύνολα με τα οποία ασχολείται η Περιγραφική Στατιστική ονομάζονται *πληθυσμοί*, και τα στοιχεία τους *στατιστικές μονάδες* ή *άτομα* ή *στοιχεία*. Οι ονομασίες αυτές προέρχονται από τη Δημογραφία, που αποτέλεσε ένα από τα πρώτα πεδία εφαρμογής της Στατιστικής, και χρησιμοποιούνται τόσο για σύνολα προσώπων όσο και για σύνολα αντικειμένων (συγκεκριμένων ή αφηρημένων).

Τα σύνολα που μελετάμε είναι απαραίτητο να ορίζονται με απόλυτη ακρίβεια. Όταν, για παράδειγμα, αναφερόμαστε στον πληθυσμό μιας πόλης, είναι αναγκαίο να προσδιορίζουμε ποιους θεωρούμε ως κατοίκους· π.χ., αν στους κατοίκους συμπεριλαμβάνουμε τους στρατιωτικούς που στρατοπεδεύουν στην πόλη, τους ασθενείς των νοσοκομείων, τους φυλακισμένους κ.ο.κ. Ανάλογα με τον τρόπο με τον οποίο ορίζουμε τον κάτοικο μιας πόλης, καθορίζουμε και διαφορετικά σύνολα, δηλαδή διαφορετικούς πληθυσμούς της συγκεκριμένης πόλης.

Οι πληθυσμοί με τους οποίους ασχολείται η Στατιστική, μπορεί καταρχήν να είναι πεπερασμένοι ή άπειροι. Στην Περιγραφική όμως Στατιστική οι πληθυσμοί που εξετάζουμε είναι πάντα πεπερασμένοι.

1.2. *Χαρακτηριστικά*

Κάθε άτομο του πληθυσμού μπορεί να εξεταστεί ως πρός ένα ή περισσότερα χαρακτηριστικά· π.χ., το προσωπικό μιας επιχείρησης (πλη-

θυσμός) μπορεί να εξεταστεί ως προς το φύλο, την ηλικία, τον μηνιαίο μισθό, την αρχαιότητα στην επιχείρηση, την ειδίκευση, τον τόπο κατοικίας κ.ο.κ. Κάθε χαρακτηριστικό μπορεί να παρουσιάζει δύο ή περισσότερες τάξεις (ή κατηγορίες). Οι τάξεις ενός χαρακτηριστικού είναι οι διάφορες καταστάσεις στις οποίες μπορεί να βρεθούν τα άτομα του πληθυσμού ως προς το χαρακτηριστικό αυτό. Οι τάξεις ενός χαρακτηριστικού είναι ασυμβίβαστες μεταξύ τους (δηλαδή δεν μπορεί να υπάρχει άτομο που να ανήκει ταυτόχρονα σε δύο τάξεις) και καλύπτουν όλες τις δυνατές καταστάσεις στις οποίες μπορεί να βρεθούν τα άτομα του πληθυσμού ως προς το χαρακτηριστικό αυτό, έτσι ώστε κάθε άτομο του πληθυσμού να ανήκει σε μία και μόνο μία τάξη του χαρακτηριστικού που εξετάζουμε. Π.χ., το χαρακτηριστικό «φύλο» παρουσιάζει δύο τάξεις — αρσενικό/θηλυκό· είναι προφανές ότι όλα τα άτομα του πληθυσμού του προηγούμενου παραδείγματος ανήκουν ή στη μία ή στην άλλη τάξη του χαρακτηριστικού «φύλο», και ότι δεν υπάρχει άτομο που ν' ανήκει ταυτόχρονα και στις δύο.

Τα χαρακτηριστικά κατατάσσονται σε δύο γενικές κατηγορίες: τα *ποιοτικά* χαρακτηριστικά και τα *ποσοτικά* χαρακτηριστικά.

1.2.1. *Ποιοτικά Χαρακτηριστικά*

Ένα χαρακτηριστικό λέγεται *ποιοτικό* όταν αφορά μια ιδιότητα των ατόμων του πληθυσμού η οποία δεν εκφράζεται με μια συγκεκριμένη μονάδα μέτρησης, όπως π.χ. το φύλο, η εθνικότητα, το επάγγελμα, η οικογενειακή κατάσταση κ.ο.κ.

Οι τάξεις ενός ποιοτικού χαρακτηριστικού είναι οι διάφορες κατηγορίες μιας ονοματολογίας, κατασκευασμένης με τέτοιο τρόπο ώστε κάθε άτομο του πληθυσμού να αντιστοιχεί σε μία και μόνο μία κατηγορία. Ένα τέτοιο παράδειγμα είναι η ονοματολογία των επαγγελματικών κατηγοριών της ΕΣΥΕ (Εθνική Στατιστική Υπηρεσία της Ελλάδας).

Συχνά τα ποιοτικά χαρακτηριστικά χωρίζονται σε δύο είδη, ανάλογα με το αν θεωρούμε ότι υπάρχει ή όχι μια ιεράρχηση μεταξύ των τάξεων του χαρακτηριστικού. Όταν μια τέτοια ιεράρχηση δεν υπάρχει, το χαρακτηριστικό λέγεται *κατηγορικό* (nominal), όπως π.χ. το φύλο, η εθνικότητα κ.ο.κ., ενώ όταν υπάρχει το χαρακτηριστικό λέγεται *τακτικό* (ordinal), όπως π.χ. το επίπεδο των γραμματικών γνώσεων. Πρέπει να σημειώσουμε ότι η ιεράρχηση μπορεί να υπάρχει είτε γιατί, κατά κάποιο τρόπο, ενυπάρχει στην ίδια τη φύση του χαρακτηριστικού (π.χ. γραμματικές γνώσεις), είτε γιατί εμείς αυθαίρετα την κατασκευάζουμε (π.χ. ιεράρχηση των επαγγελματικών κατηγοριών από την ΕΣΥΕ). Η διάκριση των ποιοτικών χαρακτηριστικών σε κατηγορικά και τακτικά εφαρμόζεται πολύ συχνά στις κοινωνικές επιστήμες, και ιδιαίτερα στην κατασκευή κλιμάκων.

1.2.2. *Ποσοτικά Χαρακτηριστικά*

Ένα χαρακτηριστικό ονομάζεται *ποσοτικό* όταν αφορά μια ιδιότητα η οποία εκφράζεται με μια συγκεκριμένη μονάδα μέτρησης. Ποσοτικά χαρακτηριστικά είναι, π.χ., η ηλικία, το βάρος, η θερμοκρασία, ο αριθμός των εργατών μιας επιχείρησης κ.ο.κ. Σ' αυτή την περίπτωση, σε κάθε τάξη του χαρακτηριστικού αντιστοιχεί ένας αριθμός (και μόνον ένας). Η αντιστοίχηση αυτή λέγεται *στατιστική μεταβλητή*, και οι αριθμοί που αντιστοιχούν στις διάφορες τάξεις του χαρακτηριστικού λέγονται *τιμές* της στατιστικής μεταβλητής.

Όταν οι τιμές μιας στατιστικής μεταβλητής είναι διακεκριμένες, δηλαδή όταν είναι μεμονωμένοι αριθμοί, η στατιστική μεταβλητή λέγεται *ασυνεχής*. Στα περισσότερα παραδείγματα, οι τιμές που παίρνουν οι ασυνεχείς μεταβλητές είναι ένα πεπερασμένο πλήθος ακέραιων αριθμών, όπως π.χ. η ηλικία ενός ατόμου εκφρασμένη σε έτη, ο αριθμός εργατών μιας επιχείρησης, ο αριθμός των μελών ενός συνεταιρισμού. Όταν, αντίθετα, μια στατιστική μεταβλητή μπορεί να πάρει οποιαδήποτε τιμή μεταξύ δύο πραγματικών αριθμών, η στατιστική μεταβλητή λέγεται *συνεχής*, όπως π.χ. η ακριβής ηλικία ενός ατόμου, η θερμοκρασία ενός σώματος, η ταχύτητα ενός κινητού.

Κατά κανόνα, συνεχείς στατιστικές μεταβλητές είναι όλα τα μεγέθη που αναφέρονται στο χώρο (μήκος, επιφάνεια, όγκος), στο χρόνο (ηλικία, διάρκεια ζωής), στη μάζα (βάρος, περιεκτικότητα), ή αποτελούν συνδυασμούς τέτοιων μεγεθών (ταχύτητα, πυκνότητα).

1.2.3. *Παρατηρήσεις*

(α΄) Από μαθηματική άποψη, ένα χαρακτηριστικό —είτε ποιοτικό είτε ποσοτικό— είναι μια συνάρτηση με πεδίο ορισμού τον πληθυσμό και πεδίο τιμών το σύνολο των τάξεων του χαρακτηριστικού. Η διάκριση των χαρακτηριστικών σε ποιοτικά (κατηγορικά ή τακτικά) και ποσοτικά γίνεται με βάση τις ιδιότητες του πεδίου τιμών (δηλαδή του συνόλου των τάξεων).

—Όταν δεν μπορούμε να συγκρίνουμε τις τάξεις μεταξύ τους, έχουμε ένα κατηγορικό χαρακτηριστικό.

—Όταν μπορούμε να τις συγκρίνουμε μεταξύ τους αλλά δεν μπορούμε να υπολογίσουμε τη διαφορά τους, έχουμε ένα τακτικό χαρακτηριστικό.

—Όταν μπορούμε να τις συγκρίνουμε μεταξύ τους και να υπολογίσουμε τη διαφορά τους, έχουμε ένα ποσοτικό χαρακτηριστικό.

(β΄) Επίσης, η στατιστική μεταβλητή, όπως την ορίσαμε πιο πάνω, είναι στην πραγματικότητα μια συνάρτηση με πεδίο ορισμού το σύνολο των τάξεων ενός ποσοτικού χαρακτηριστικού και πεδίο τιμών το σύνολο των αριθμών που αντιστοιχούν στις διάφορες τάξεις.[1]

Η διάκριση μεταξύ συνεχών και ασυνεχών στατιστικών μεταβλητών αφορά λοιπόν το πεδίο τιμών της μεταβλητής. Όμως η αντιστοίχηση ενός αριθμού σε μια τάξη του ποσοτικού χαρακτηριστικού γίνεται με βάση μια μέτρηση, και κάθε μέτρηση είναι ασυνεχής γιατί η ακρίβειά της είναι πάντα περιορισμένη. Για το λόγο αυτό, μερικές φορές θεωρούμε ως ασυνεχείς κάποιες στατιστικές μεταβλητές που στην πραγματικότητα είναι συνεχείς. Αντίστροφα, πολλές φορές θεωρούμε ότι εκφράζονται με συνεχείς μεταβλητές μεγέθη που είναι δυνατό να πάρουν έναν μεγάλο αριθμό τιμών ακόμα κι αν στην πραγματικότητα οι τιμές αυτές είναι πεπερασμένες. Έτσι, π.χ., τα χρηματικά μεγέθη (μισθός, εισόδημα, κέρδος) θεωρούνται ως ασυνεχείς μεταβλητές παρόλο που οι τιμές μιας τέτοιας μεταβλητής είναι διακεκριμένες, και μάλιστα πεπερασμένες.

1. Θα έπρεπε ίσως ως στατιστική μεταβλητή να ορίσουμε την εξαρτημένη μεταβλητή αυτής της συνάρτησης. Το αποφύγαμε, ελπίζοντας ότι έτσι απλοποιείται η παρουσίαση των βασικών εννοιών.

ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ

ΠΟΙΟΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

Η μελέτη ενός πληθυσμού ως προς ένα ποιοτικό χαρακτηριστικό δεν διαφέρει ουσιαστικά από τη μελέτη του ως προς ένα ποσοτικό χαρακτηριστικό. Είναι όμως προτιμότερο να παρουσιαστούν τα δύο θέματα χωριστά, ώστε ν' αποκτήσει ευκολότερα ο αναγνώστης οικειότητα με τις κυριότερες τεχνικές της Περιγραφικής Στατιστικής, κι επίσης γιατί πολύ συχνά στις κοινωνικές επιστήμες θα έχει να κάνει με ποιοτικά χαρακτηριστικά.

2.1. *Στατιστικοί Πίνακες*

Ας θεωρήσουμε έναν πληθυσμό Π, που αποτελείται απο ν άτομα (A_1, A_2, ..., A_ν), κι ένα ποιοτικό χαρακτηριστικό X, που έχει κ τάξεις (X_1, ..., X_κ). Επειδή οι τάξεις του χαρακτηριστικού είναι ασυμβίβαστες μεταξύ τους και καλύπτουν όλες τις δυνατές καταστάσεις (βλ. §1.2.), καθένα από τα ν άτομα του πληθυσμού αντιστοιχεί σε μία και μόνο μία τάξη του χαρακτηριστικού.

Τον αριθμό των ατόμων που αντιστοιχούν σε μία τάξη του χαρακτηριστικού τον ονομάζουμε *απόλυτη συχνότητα* αυτής της τάξης. Το πρώτο βήμα για τη μελέτη ενός πληθυσμού ως προς ένα χαρακτηριστικό είναι να βρούμε πόσα άτομα του πληθυσμού αντιστοιχούν σε κάθε μία από τις τάξεις του, δηλαδή την απόλυτη συχνότητα της κάθε τάξης. Ονομάζουμε *στατιστικό πίνακα* του πληθυσμού Π ως προς το χαρακτηριστικό X του πίνακα που μας δίνει για κάθε μία τάξη του χαρακτηριστικού την απόλυτη συχνότητά της. Η γενική μορφή ενός τέτοιου στατιστικού πίνακα είναι η εξής:

ΠΙΝΑΚΑΣ 2.1.

Τάξεις του χαρακτηριστικού Χ	*Απόλυτη συχνότητα κάθε τάξης*
X_1	ν_1
X_2	ν_2
⋮	⋮
X_i	ν_i
⋮	⋮
X_κ	ν_κ
Σύνολο	ν

Οι στατιστικοί πίνακες ενός πληθυσμού ως προς ένα χαρακτηριστικό ονομάζονται επίσης και *πίνακες απλής εισόδου*.

Με την κατασκευή του στατιστικού πίνακα περνάμε σ' ένα πρώτο επίπεδο ανωνυμίας. Πραγματικά, ενώ όταν αντιστοιχίζουμε σε κάθε άτομο μια τάξη του χαρακτηριστικού το κάθε άτομο θεωρείται διαφορετικό από τα υπόλοιπα, όταν κατασκευάσουμε τον στατιστικό πίνακα όλα τα άτομα που ανήκουν σε μια συγκεκριμένη τάξη του χαρακτηριστικού θεωρούνται ισοδύναμα μεταξύ τους, δηλαδή ταυτίζονται ως προς το χαρακτηριστικό που εξετάζουμε και μας είναι γνωστά μέσω του χαρακτηριστικού αυτού.

2.2. *Σχετική Συχνότητα*

Αν ν_i είναι η απόλυτη συχνότητα της τάξης X_i και ν ο αριθμός των ατόμων του πληθυσμού, το πηλίκο:

$$f_i = \frac{\nu_i}{\nu}$$

ονομάζεται *σχετική συχνότητα* ή *αναλογία* της τάξης X_i. Η σχετική συχνότητα είναι ένας δεκαδικός αριθμός μεταξύ 0 και 1. Είναι ίση με το

0 όταν κανένα άτομο του πληθυσμού δεν αντιστοιχεί στη συγκεκριμένη τάξη, και ίση με το 1 όταν όλα τα άτομα του πληθυσμού αντιστοιχούν στη συγκεκριμένη τάξη.

Μια και το άθροισμα των απόλυτων συχνοτήτων (ν_i) είναι ίσο με τον συνολικό αριθμό ν των ατόμων του πληθυσμού, το άθροισμα των σχετικών συχνοτήτων είναι ίσο με τη μονάδα:

$$\sum_{i=1}^{\kappa} \nu_i = \nu \quad \text{και} \quad \sum_{i=1}^{\kappa} f_i = \sum_{i=1}^{\kappa} \frac{\nu_i}{\nu} = \frac{1}{\nu} \sum_{i=1}^{\kappa} \nu_i = 1.$$

Οι τάξεις του χαρακτηριστικού μαζί με τις αντίστοιχες σχετικές συχνότητες αποτελούν αυτό που ονομάζουμε *κατανομή συχνοτήτων* (στη συγκεκριμένη περίπτωση ως προς ένα ποιοτικό χαρακτηριστικό).

Συχνά, αντί για τις σχετικές συχνότητες χρησιμοποιούμε τις αναλογίες επί τοις εκατό ή ποσοστά επί τοις εκατό (%). Τα ποσοστά επί τοις εκατό είναι ίσα με τις σχετικές συχνότητες πολλαπλασιασμένες επί 100, και δείχνουν πόσα άτομα θ' αντιστοιχούσαν στις διάφορες τάξεις του χαρακτηριστικού αν το συνολικό μέγεθος του πληθυσμού ήταν ίσο με 100. Μερικές φορές χρησιμοποιούμε επίσης και τα ποσοστά επί τοις χιλίοις (‰), που δείχνουν πόσα άτομα του πληθυσμού θ' αντιστοιχούσαν στις διάφορες τάξεις του χαρακτηριστικού αν το συνολικό μέγεθος του πληθυσμού ήταν ίσο με 1.000.

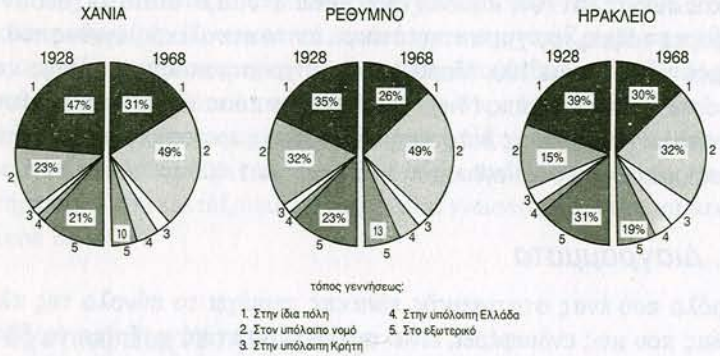
2.3. *Διαγράμματα*

Παρόλο που ένας στατιστικός πίνακας περιέχει το σύνολο της πληροφορίας που μας ενδιαφέρει, είναι συχνά εξαιρετικά χρήσιμο να δίνεται, μ' ένα ή περισσότερα διαγράμματα, μια οπτική εικόνα του πίνακα. Η κατασκευή διαγραμμάτων είναι χρήσιμη όχι μόνο για την παρουσίαση των αποτελεσμάτων, αλλά και κατά τη διάρκεια της μελέτης ενός φαινομένου. Πραγματικά, η μεταφορά των αριθμητικών δεδομένων σε γεωμετρικά σχήματα, όταν γίνεται με σωστό τρόπο, είναι δυνατόν να βοηθήσει τον ερευνητή να εντοπίσει μερικές όψεις του φαινομένου που δύσκολα θα μπορούσε ν' αντιληφθεί μελετώντας μόνο τον στατιστικό πίνακα.

Ανάλογα με τη φύση του χαρακτηριστικού που μελετάμε, χρησιμοποιούμε διάφορα είδη γραφικών παραστάσεων. Για τη γραφική παράσταση των κατανομών ως προς ένα ποιοτικό χαρακτηριστικό χρησιμοποιούμε κατά κανόνα τα *κυκλικά* και *ημικυκλικά διαγράμματα*, καθώς και τα *ακιδωτά διαγράμματα* (ή *ραβδογράμματα*).

Στα κυκλικά και τα ημικυκλικά διαγράμματα, σε κάθε τάξη του χαρακτηριστικού αντιστοιχεί ένας κυκλικός τομέας που την παριστάνει, και που το εμβαδόν του είναι ανάλογο προς την απόλυτη συχνότητα της τάξης. Επομένως, αν X_i είναι μια τάξη του χαρακτηριστικού και f_i η σχετική της συχνότητα, το τόξο του κυκλικού τομέα που την παριστάνει ισούται με $360° \times f_i$ για τα κυκλικά διαγράμματα και με $180° \times f_i$ για τα ημικυκλικά. Π.χ.:

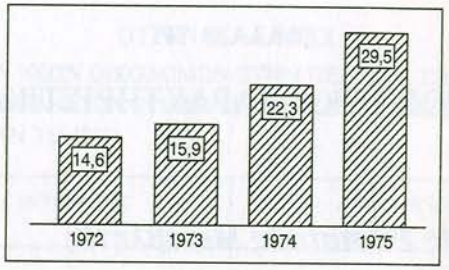
ΣΧΗΜΑ 2.1. *Περιοχές προέλευσης των κατοίκων των πόλεων της Κρήτης το 1928 και το 1967-68 (τόπος γεννήσεως του ενήλικου πληθυσμού).*



ΠΗΓΗ: EMILE Y. KOLODNY, *La population des iles de la Grèce* (εκδ. Edisud, Aix-en-Provence, 1974, τ. 3, Διάγρ. J.18).

Τα ακιδωτά διαγράμματα αποτελούνται από ορθογώνια παραλληλόγραμμα που έχουν ίσες βάσεις και ύψη ανάλογα προς την απόλυτη συχνότητα της κάθε τάξης του χαρακτηριστικού. Π.χ.:

ΣΧΗΜΑ 2.2. *Έλλειμμα του γενικού κρατικού προϋπολογισμού (σε δισ. δραχμές σε τρέχουσες τιμές).*



ΠΗΓΗ: Η ελληνική βιομηχανία κατά το 1975 (έκδ. Συνδέσμου Ελλήνων Βιομηχάνων, Αθήνα, 1976).

Γενικός κανόνας τόσο για τα κυκλικά όσο και για τα ακιδωτά διαγράμματα, καθώς και για οποιαδήποτε άλλη γραφική παράσταση κατανομής ενός πληθυσμού ως προς ένα ποιοτικό χαρακτηριστικό, είναι ότι τα εμβαδά πρέπει να είναι ανάλογα προς τις απόλυτες συχνότητες (ή, πράγμα που για τα ποιοτικά χαρακτηριστικά είναι ισοδύναμο, να είναι ανάλογα προς τις σχετικές συχνότητες).

ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ

ΠΟΣΟΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

3.1. *Ασυνεχείς Στατιστικές Μεταβλητές*

Όπως και στην περίπτωση των ποιοτικών χαρακτηριστικών, έτσι και για τα ποσοτικά χαρακτηριστικά που εκφράζονται με ασυνεχείς στατιστικές μεταβλητές ο στατιστικός πίνακας μας δίνει για κάθε τάξη του χαρακτηριστικού (X), κι επομένως για κάθε τιμή (X_i) της στατιστικής μεταβλητής, τον αριθμό των ατόμων που αντιστοιχούν σ' αυτή την τάξη, δηλαδή την απόλυτη συχνότητά της (ν_i). Η γενική μορφή ενός τέτοιου στατιστικού πίνακα, όταν οι τιμές της στατιστικής μεταβλητής είναι πεπερασμένες, είναι η εξής:

ΠΙΝΑΚΑΣ 3.1.

Τιμή της μεταβλητής	*Απόλυτη συχνότητα*
X_1	ν_1
⋮	⋮
X_i	ν_i
⋮	⋮
X_κ	ν_κ
Σύνολο	ν

Μερικές φορές συγκεντρώνουμε σε μια κατηγορία τις ακραίες τιμές της μεταβλητής. Αυτό γίνεται κατά κανόνα όταν οι αντίστοιχες σχετικές συχνότητες είναι πολύ μικρές. Π.χ.:

ΠΙΝΑΚΑΣ 3.2.

ΚΑΤΑΝΟΜΗ ΤΩΝ ΝΕΩΝ ΟΙΚΟΔΟΜΩΝ ΣΤΗΝ ΠΕΡΙΟΧΗ ΤΗΣ ΠΡΩΤΕΥΟΥΣΑΣ ΩΣ ΠΡΟΣ ΤΟΝ ΑΡΙΘΜΟ ΤΩΝ ΟΡΟΦΩΝ (ΜΕ ΒΑΣΗ ΤΙΣ ΟΙΚΟΔΟΜΙΚΕΣ ΑΔΕΙΕΣ ΠΟΥ ΕΚΔΟΘΗΚΑΝ ΤΟ 1973).

Τιμές της μεταβλητής	*Απόλυτη συχνότητα (αριθμός οικοδομών)*
1 όροφος	3.904
2 όροφοι	3.103
3 »	1.208
4 »	819
5 »	836
6 »	1.226
7 »	280
8 »	74
9 »	26
10 » ή περισσότεροι	11
Σύνολο	11.487

ΠΗΓΗ: Στατιστική Επετηρίς της Ελλάδος (1974, σ. 262).

Σε κάθε τιμή της στατιστικής μεταβλητής (X_i) αντιστοιχεί και μια σχετική συχνότητα (f_i), που ορίζεται, όπως και στην περίπτωση των ποιοτικών χαρακτηριστικών, ως το πηλίκο της απόλυτης συχνότητας (ν_i) προς το σύνολο του πληθυσμού (ν), δηλαδή:

$$f_i = \frac{\nu_i}{\nu}.$$

Το άθροισμα των σχετικών συχνοτήτων είναι ίσο με τη μονάδα, δηλαδή:

$$\sum_{i=1}^{\kappa} f_i = 1 \quad (\text{βλ. } §2.2.).$$

Οι τιμές της στατιστικής μεταβλητής και οι αντίστοιχες σχετικές συχνότητες ορίζουν, όπως και προηγουμένως, την κατανομή των συχνοτήτων.

Εκτός από τις σχετικές συχνότητες (f_i), οι οποίες ορίζονται για τις διάφορες τιμές (x_i) της στατιστικής μεταβλητής, ορίζουμε επίσης για κάθε πραγματικό αριθμό x την ολική σχετική συχνότητα του x. Η **ολική σχετική συχνότητα** του x είναι το πηλίκο του αριθμού των ατόμων του πληθυσμού για τα οποία η τιμή της στατιστικής μεταβλητής είναι **μικρότερη** από x, ως προς το σύνολο του πληθυσμού. Αν τώρα, σε κάθε πραγματικό αριθμό x, αντιστοιχίσουμε την ολική σχετική συχνότητα του x, ορίζουμε μια συνάρτηση που λέγεται *αθροιστική συνάρτηση* και συμβολίζεται με $F(x)$.

Η αθροιστική συνάρτηση $F(x)$:

— είναι σταθερή στο διάστημα που ορίζουν δύο διαδοχικές τιμές της στατιστικής μεταβλητής, δηλαδή:

$$F(x)=\sum_{\lambda=1}^{i} fx \text{ για κάθε } x: x_i < x \leqslant x_{i+1},$$

— παρουσιάζει σε κάθε τιμή της στατιστικής μεταβλητής ένα **άλμα** ίσο με την αντίστοιχη σχετική συχνότητα, δηλαδή:

$$F(x_i)=\sum_{j=1}^{i-1} f_j, \qquad F(x_i+0)=\sum_{i=1}^{i} f_j,$$

κι επομένως: $$F(x_i+0)-F(x_i)=f_i,$$

— είναι ίση με το μηδέν για κάθε x **μικρότερο ή ίσο** από τη μικρότερη τιμή της στατιστικής μεταβλητής και ίση με το 1 για κάθε x μεγαλύτερο από τη μεγαλύτερη τιμή της στατιστικής μεταβλητής.

Όταν έχουμε έναν στατιστικό πίνακα, για να βρούμε τις τιμές της αθροιστικής συνάρτησης $F(x)$, δηλαδή τις ολικές σχετικές συχνότητες, προσθέτουμε διαδοχικά τις σχετικές συχνότητες f_i. Τις τιμές που βρίσκουμε, τις γράφουμε **μεταξύ** των γραμμών του πίνακα, για να δείξουμε ότι αναφέρονται στο διάστημα που χωρίζει δύο διαδοχικές τιμές της στατιστικής μεταβλητής. Π.χ., ο πίνακας του προηγούμενου παραδείγματος παίρνει την εξής μορφή αν τον συμπληρώσουμε με τις σχετικές συχνότητες και τις ολικές σχετικές συχνότητες:

ΠΙΝΑΚΑΣ 3.3.

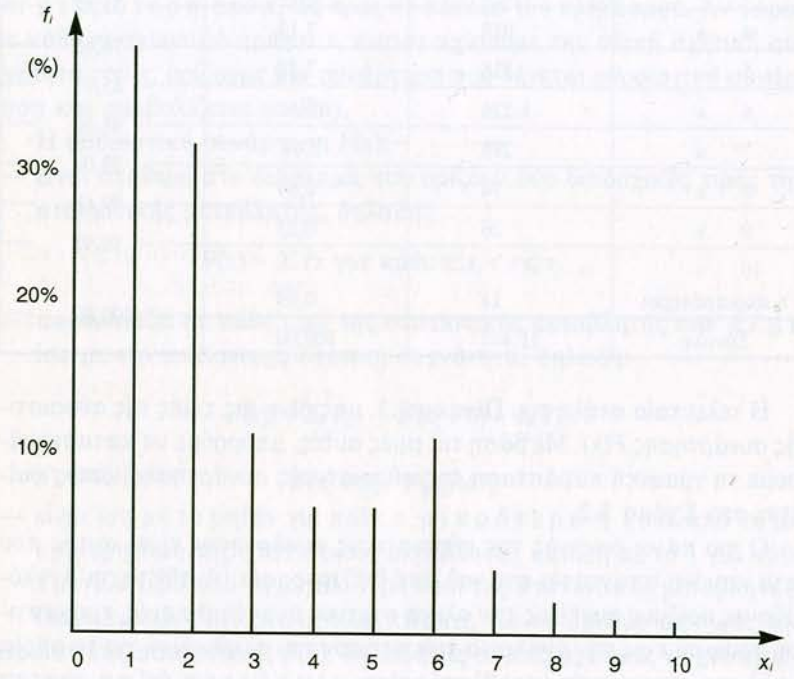
Τιμές της μεταβλητής x_i	*Απόλυτες συχνότητες* (αριθμός οικοδομών) v_i	*Σχετικές συχνότητες* (σε %) f_i	*Ολικές σχετικές συχνότητες* (σε %) $F(x_i)$
1 όροφος	3.904	33,99	33,99
2 όροφοι	3.103	27,01	61,00
3 »	1.208	10,52	71,52
4 »	819	7,13	78,65
5 »	836	7,28	85,93
6 »	1.226	10,67	96,60
7 »	280	2,44	99,04
8 »	74	0,64	99,68
9 »	26	0,23	99,91
10 » ή περισσότεροι	11	0,09	100,00
Σύνολο	11.487	100,00	

Η τελευταία στήλη του Πίνακα 3.3. μας δίνει τις τιμές της αθροιστικής συνάρτησης $F(x)$. Με βάση τις τιμές αυτές, μπορούμε να κατασκευάσουμε τη γραφική παράσταση της αθροιστικής συνάρτησης, όπως φαίνεται στο Σχήμα 3.2.

Ο πιο πάνω ορισμός της αθροιστικής συνάρτησης είναι αυτός που κατά κανόνα απαντάται στη γαλλική βιβλιογραφία. Αντίθετα, οι Αγγλοσάξονες ορίζουν συνήθως την ολική σχετική συχνότητα ενός πραγματικού αριθμού x ως την αναλογία των ατόμων του πληθυσμού για τα οποία η τιμή της στατιστικής μεταβλητής είναι **μικρότερη ή ίση** προς το x. Η διαφορά των δύο ορισμών της αθροιστικής συνάρτησης αφορά μόνο τα x που είναι τιμές της στατιστικής μεταβλητής. Δηλαδή, κατά τον γαλλικό ορισμό: $F(x_i)=f_1+\ldots+f_{i-1}$, ενώ κατά τον αγγλοσαξονικό: $F(x_i)=f_1+\ldots+f_i$.

Για τη γραφική παράσταση ενός ποσοτικού χαρακτηριστικού που εκφράζεται με ασυνεχή στατιστική μεταβλητή χρησιμοποιούνται δύο ειδών διαγράμματα: το *ιστόγραμμα*, το οποίο απεικονίζει τις απόλυτες συχνότητες ν_i (ή, πράγμα ισοδύναμο, τις σχετικές συχνότητες f_i) που αντιστοιχούν στις διάφορες τιμές x_i της στατιστικής μεταβλητής, και η *αθροιστική καμπύλη*, όπως ονομάζεται η γραφική παράσταση της αθροιστικής συνάρτησης (βλ. Σχ. 3.1. και 3.2.).

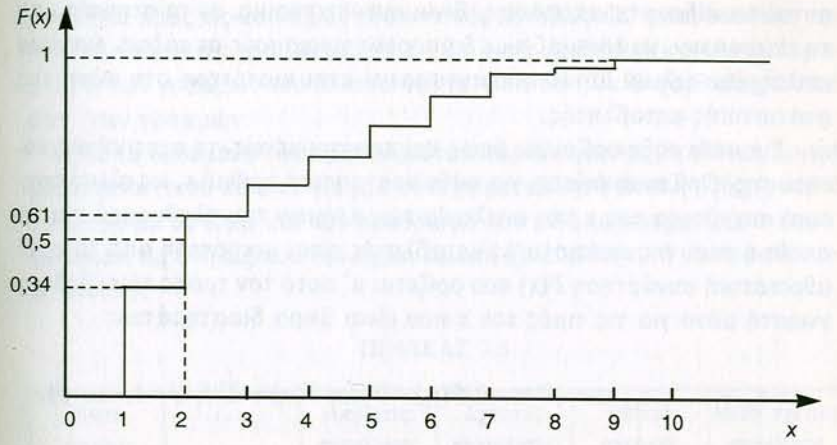
ΣΧΗΜΑ 3.1. *Ιστόγραμμα της κατανομής των νέων οικοδομών (1973).*



3.2. *Συνεχείς Στατιστικές Μεταβλητές*

Όταν το ποσοτικό χαρακτηριστικό εκφράζεται με μία συνεχή στατιστική μεταβλητή, είναι φυσικά αδύνατο ο στατιστικός πίνακας να δίνει για κάθε δυνατή τιμή της στατιστικής μεταβλητής την απόλυτη συχνότητά

ΣΧΗΜΑ 3.2. *Αθροιστική συνάρτηση της κατανομής των νέων οικοδομών (1973).*



της. Γι' αυτό, ως τάξεις του χαρακτηριστικού ορίζουμε διαστήματα τιμών της στατιστικής μεταβλητής, και η κάθε τάξη προσδιορίζεται από τις ακραίες τιμές τους. Γενικά, αν συμβολίσουμε με e_0, e_1, e_2, ..., e_i, ..., e_κ τις ακραίες τιμές (άκρα) των διαφόρων διαστημάτων που θεωρούμε, η τάξη με αρίθμηση i περιλαμβάνει όλες τις τιμές της στατιστικής μεταβλητής που είναι μικρότερες από το e_i και μεγαλύτερες ή ίσες από το e_{i-1}, *δηλαδή* $e_{i-1} \leqslant x < e$, $(i = 1, \ldots, \kappa)$.

Κέντρο της τάξης που αριθμείται με i ονομάζεται ο αριθμός:

$$C_i = \frac{e_{i-1} + e_i}{2}.$$

Η απόσταση ανάμεσα στα κέντρα των τάξεων που αριθμούνται με i και $i+1$ είναι ίση με:

$$d_i = C_{i+1} - C_i = \frac{e_{i+1} - e_{i-1}}{2}.$$

Το *πλάτος* της τάξης με αρίθμηση i είναι ίσο με τη διαφορά των άκρων της, δηλαδή $a_i = e_i - e_{i-1}$.

Όταν πρόκειται να κατασκευάσουμε τον στατιστικό πίνακα ενός πληθυσμού ως προς μια συνεχή στατιστική μεταβλητή, αντιμετωπίζουμε πάντοτε το πρόβλημα του αριθμού των τάξεων που θα ορίσουμε και του κα-

θορισμού των ακραίων τιμών τους. Δυστυχώς δεν υπάρχουν κανόνες γι' αυτού του είδους τις αποφάσεις. Είναι συχνά χρήσιμο, αν τα στοιχεία μάς το επιτρέπουν, να δοκιμάζουμε διάφορους χωρισμούς σε τάξεις, ώστε να επιλέξουμε τελικά αυτόν που ανταποκρίνεται πιστότερα στη φύση της στατιστικής μεταβλητής.

Για κάθε τάξη ορίζουμε, όπως και προηγουμένως, τη σχετική συχνότητά της. Ορίζουμε επίσης, για κάθε πραγματικό αριθμό x, ως ολική σχετική συχνότητα του x την αναλογία των ατόμων του πληθυσμού για τα οποία η τιμή της στατιστικής μεταβλητής είναι μικρότερη από το x. Η αθροιστική συνάρτηση $F(x)$ που ορίζεται μ' αυτό τον τρόπο είναι βέβαια γνωστή μόνο για τις τιμές του x που είναι άκρα διαστημάτων:

$$x = e_0, e_1, \ldots, e_\kappa, \text{ και γι' αυτά } F(e_0) = 0 \text{ και } F(e_i) = \sum_{\lambda=1}^{i} f_\lambda \ (i = 1, \ldots, \kappa).$$

Η πιο συνηθισμένη μορφή με την οποία συναντάμε τους στατιστικούς πίνακες των συνεχών στατιστικών μεταβλητών είναι η εξής:

ΠΙΝΑΚΑΣ 3.4.

ΚΑΤΑΝΟΜΗ ΤΩΝ ΑΓΡΟΤΙΚΩΝ ΕΚΜΕΤΑΛΛΕΥΣΕΩΝ ΩΣ ΠΡΟΣ ΤΟ ΜΕΓΕΘΟΣ (ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ 5% ΤΩΝ ΔΕΛΤΙΩΝ ΤΗΣ ΑΠΟΓΡΑΦΗΣ ΓΕΩΡΓΙΑΣ-ΚΤΗΝΟΤΡΟΦΙΑΣ 1971).

Μέγεθος αγροτικών εκμεταλλεύσεων (εκτάσεις σε στρέμματα)	*Αριθμός εκμεταλλεύσεων*
0 - 9	225.820
10 - 29	384.320
30 - 49	209.640
50 - 99	164.340
100 - 199	42.760
200 - 499	8.840
500 και άνω	880
Σύνολο	1.036.600

ΠΗΓΗ: Στατιστική Επετηρίς της Ελλάδος (1974, σ. 158).

Είναι όμως προτιμότερο, για την ευκολία των υπολογισμών και από αναλογία προς τις ασυνεχείς στατιστικές μεταβλητές, η παρουσίαση του πίνακα να γίνεται γράφοντας τα ε ν δ ο τ α ξ ι κ ά χαρακτηριστικά μ ε τ α ξ ύ των γραμμών του πίνακα και τα δ ι α τ α ξ ι κ ά χαρακτηριστικά ε π ί των γραμμών.

Με τα δεδομένα του προηγούμενου παραδείγματος, η γενική μορφή του στατιστικού πίνακα για μια συνεχή μεταβλητή είναι η εξής (φυσικά ανάλογα με το θέμα και τον προορισμό του πίνακα μπορούμε να παραλείπουμε τις στήλες που δεν έχουν άμεση χρησιμότητα ή να προσθέτουμε άλλες):

ΠΙΝΑΚΑΣ 3.5.

Άκρα τάξεων (εκτάσεις σε στρέμματα) e_i	*Πλάτος* a_i	*Απόλυτες συχνότητες (αριθμός εκμεταλλεύσεων)* v_i	*Σχετικές συχνότητες (σε %)* f_i	*Ολικές σχετικές συχνότητες (σε %)* $F(e_i)$	*Μέση σχετική συχνότητα ανά μονάδα πλάτους* f_i/a_i
10	10	225.820	21,78	21,78	2,178
30	20	384.320	37,08	58,86	1,854
50	20	209.640	20,22	71,08	1,011
100	50	164.340	15,85	94,93	0,317
200	100	42.760	4,13	99,06	0,041
500	300	8.840	0,85	99,91	0,003
	απροσδιόριστο	880	0,09	100,00	απροσδιόριστη
Σύνολο	—	1.036.600	100,00	—	—

Όπως συμβαίνει και με τις ασυνεχείς στατιστικές μεταβλητές, συχνά οι ακραίες τάξεις προσδιορίζονται με εκφράσεις όπως «μικρότερο από...» ή «μεγαλύτερο από...», κι επομένως το πλάτος τους είναι απροσδιόριστο. Η απροσδιοριστία όμως αυτή δεν δημιουργεί σημαντικά προβλήματα αν οι αντίστοιχες σχετικές συχνότητες είναι μικρές.

Για τη γραφική παράσταση των συνεχών στατιστικών μεταβλητών χρησιμοποιούμε το *ιστόγραμμα*, την *πολυγωνική γραμμή συχνότητας* και την *αθροιστική καμπύλη*.

Το ιστόγραμμα απεικονίζει τις απόλυτες συχνότητες v που αντιστοιχούν στις διάφορες τάξεις τιμών της στατιστικής μεταβλητής. Το ιστόγραμμα εξασφαλίζει την αναλογία των εμβαδών προς τις απόλυτες συχνότητες. Αυτό συμαίνει ότι:

— όταν όλες οι τάξεις έχουν το ίδιο πλάτος, τότε το ύψος του ορθογωνίου που αντιστοιχεί σε κάθε τάξη είναι ανάλογο προς τη σχετική συχνότητα·

— όταν όμως το πλάτος δεν είναι το ίδιο για όλες τις τάξεις, τότε το ύψος του ορθογωνίου που αντιστοιχεί σε κάθε τάξη δεν είναι ανάλογο προς τη σχετική συχνότητα, αλλά είναι ανάλογο προς τη μέση σχετική συχνότητα ανά μονάδα πλάτους, δηλαδή προς τον αριθμό f_i/a_i.

Επομένως, όταν το πλάτος δεν είναι το ίδιο για όλες τις τάξεις, χρειάζεται να δώσουμε ιδιαίτερη προσοχή για να κατασκευάσουμε το ιστόγραμμα, γιατί τα λάθη είναι πολύ συχνά.

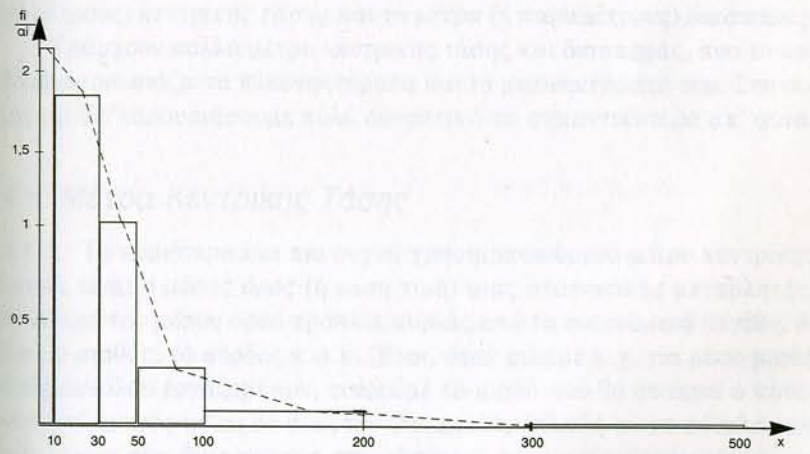
Όταν το πλάτος των ακραίων τάξεων δεν είναι προσδιορισμένο (π.χ. τάξεις που ορίζονται από εκφράσεις όπως «από 500 στρ. και άνω»), το καθορίζουμε συμβατικά. Αν π.χ. μπορούμε να υπολογίσουμε τον μέσο όρο της τάξης (βλ. §4.), θεωρούμε ως κέντρο της τάξης έναν στρογγυλεμένο αριθμό κοντά στο μέσο όρο, κι επομένως το πλάτος της τάξης είναι το διπλάσιο της απόστασης του κέντρου (που ορίσαμε συμβατικά) από το γνωστό άκρο της τάξης. Π.χ., για να υπολογίσουμε το πλάτος της τάξης «από 500 στρ. και άνω», χρησιμοποιούμε το επιπλέον στοιχείο ότι η συνολική έκταση των εκμεταλλεύσεων αυτής της κατηγορίας είναι 889.060 στρ. Άρα ο μέσος όρος (βλ. §4.1.) αυτής της τάξης είναι περίπου 1.000 στρ., κι επομένως το πλάτος αυτής της τάξης το θεωρούμε ίσο με 1.000, δηλαδή θεωρούμε ως άκρα της το 500 και το 1.500. Όμως, στη συγκεκριμένη περίπτωση, η τάξη αυτή, λόγω της εξαιρετικά μικρής σχετικής συχνότητας ανά μονάδα πλάτους, είναι πρακτικά αδύνατο να παρασταθεί στο ιστόγραμμα του Σχήματος 3.3.

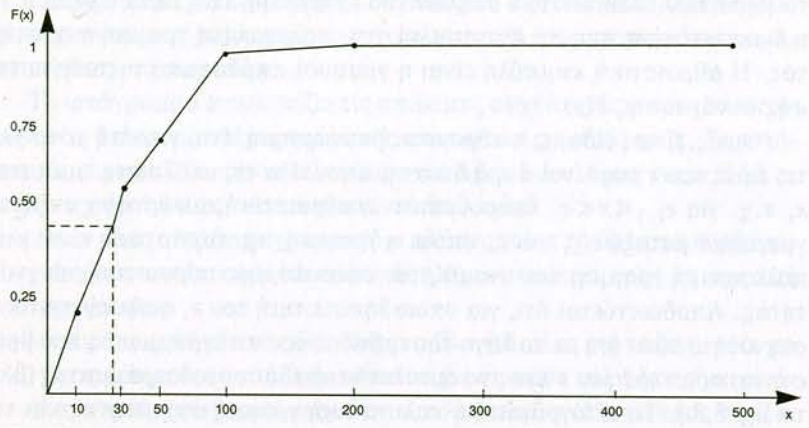
Η πολυγωνική γραμμή συχνότητας είναι η τεθλασμένη που ενώνει

τα μέσα των πλευρών των στηλών του ιστογράμματος. Στο Σχήμα 3.3. η διακεκομμένη γραμμή αντιστοιχεί στην πολυγωνική γραμμή συχνότητας. Η αθροιστική καμπύλη είναι η γραφική παράσταση της αθροιστικής συνάρτησης $F(x)$.

Όμως, όπως είδαμε, η αθροιστική συνάρτηση είναι γνωστή μόνο για τις τιμές του x που είναι άκρα διαστημάτων. Για τις υπόλοιπες τιμές του x, π.χ. για $e_{i-1}<x<e_i$, θεωρούμε ότι η αθροιστική συνάρτηση αυξάνει γραμμικά μεταξύ e_{i-1} και e_i, οπότε η γραφική της παράσταση είναι μια πολυγωνική γραμμή, που ονομάζεται και *πολύγωνο αθροιστικής συχνότητας*. Αποδεικνύεται ότι, για οποιαδήποτε τιμή του x, η ολική σχετική συχνότητα είναι ίση με το λόγο του εμβαδού του ιστογράμματος που βρίσκεται αριστερά του x προς το συνολικό εμβαδό του ιστογράμματος (βλ. το Σχ. 3.3.). Το ιστόγραμμα, η πολυγωνική γραμμή συχνότητας και το πολύγωνο αθροιστικής συχνότητας που αντιστοιχούν στα δεδομένα του στατιστικού πίνακα του προηγούμενου παραδείγματος, είναι τα εξής:

ΣΧΗΜΑ 3.3.





ΚΕΦΑΛΑΙΟ ΤΕΤΑΡΤΟ

ΑΡΙΘΜΗΤΙΚΗ ΠΕΡΙΓΡΑΦΗ ΜΙΑΣ ΣΤΑΤΙΣΤΙΚΗΣ ΜΕΤΑΒΛΗΤΗΣ

Όταν μελετάμε μια στατιστική μεταβλητή, είναι ιδιαίτερα χρήσιμο να γνωρίζουμε και να μπορούμε να συνοψίσουμε με μερικούς αριθμούς την κεντρική τάση των τιμών της μεταβλητής —δηλαδή το σημείο γύρω από το οποίο συγκεντρώνονται, κατά κάποια έννοια, αυτές οι τιμές—, καθώς και τη διασπορά τους — δηλαδή τη σχετική ομοιογένεια ή ανομοιογένειά τους. Για τον σκοπό αυτό χρησιμοποιούμε τα *μέτρα* (ή παραμέτρους) *κεντρικής τάσης* και τα *μέτρα* (ή παραμέτρους) *διασποράς*.

Υπάρχουν πολλά μέτρα κεντρικής τάσης και διασποράς, που το καθένα παρουσιάζει τα πλεονεκτήματα και τα μειονεκτήματά του. Στη συνέχεια θα παρουσιάσουμε πολύ συνοπτικά τα σημαντικότερα απ' αυτά.

4.1. *Μέτρα Κεντρικής Τάσης*

4.1.1. Το κυριότερο και πιο συχνά χρησιμοποιούμενο μέτρο κεντρικής τάσης είναι ο *μέσος όρος* (ή μέση τιμή) μιας στατιστικής μεταβλητής. Η έννοια του μέσου όρου προήλθε κυρίως από τα οικονομικά μεγέθη, όπως ο μισθός, το κέρδος κ.ο.κ. Έτσι, όταν μιλάμε π.χ. για μέσο μισθό ενός συνόλου εργαζομένων, εννοούμε το μισθό που θα έπαιρνε ο καθένας απ' αυτούς αν είχαν όλοι τον ίδιο μισθό, δηλαδή αν το σύνολο των χρημάτων που δίνονται για την πληρωμή τους μοιραζόταν εξίσου σε όλους.

Μέσος όρος ($\bar{x}$) μιας στατιστικής μεταβλητής X ονομάζεται το άθροισμα, για όλα τα άτομα του πληθυσμού, των τιμών που παίρνει η στα-

τιστική μεταβλητή διά του μεγέθους του πληθυσμού (ν), ή, πράγμα ισοδύναμο, ο σταθμισμένος (με την αντίστοιχη απόλυτη συχνότητα) αριθμητικός μέσος των τιμών της στατιστικής μεταβλητής, δηλαδή:

$$\bar{x} = \frac{1}{\nu}\sum_{\alpha=1}^{\nu} x_\alpha = \frac{1}{\nu}\sum_{i=1}^{\kappa} \nu_i x_i = \sum_{i=1}^{\kappa} f_i x_i \qquad (1)$$

Ο μέσος όρος λέγεται και *σταθμικός μέσος όρος*, γιατί κάθε τιμή x_i της στατιστικής μεταβλητής έχει διαφορετικό «βάρος», ανάλογο προς την αντίστοιχη σχετική συχνότητα f_i.

Ο μέσος όρος που ορίζεται από τη σχέση (1) λέγεται επίσης και *αριθμητικός* (ή *σταθμισμένος αριθμητικός*) *μέσος όρος*, σε αντιδιαστολή με τον *γεωμετρικό* ή τον *αρμονικό μέσο όρο*, με τους οποίους όμως δεν θα ασχοληθούμε εδώ.

Ο μέσος όρος είναι η περισσότερο χρησιμοποιούμενη στατιστική παράμετρος. Το βασικότερο ίσως μειονέκτημά του είναι ότι, στην περίπτωση των ασυνεχών μεταβλητών, δεν συμπίπτει αναγκαστικά με μια τιμή της μεταβλητής, όπως συμβαίνει με τη διάμεσο ή τον τύπο (βλ. 4.1.2. και 4.1.3.).

Στην περίπτωση των ασυνεχών στατιστικών μεταβλητών, ο υπολογισμός του μέσου όρου με βάση τη σχέση $\bar{x} = \frac{1}{\nu}\sum_{i=1}^{\kappa} \nu_i x_i$ ή τη σχέση $\bar{x} = \sum_{i=1}^{\kappa} f_i x_i$ δεν παρουσιάζει, φυσικά, καμιά δυσκολία. Όμως οι προηγούμενες σχέσεις δεν μπορούν να εφαρμοστούν στην περίπτωση των συνεχών μεταβλητών, όταν δεν γνωρίζουμε τις διάφορες τιμές της μεταβλητής αλλά μόνο τον αριθμό ν των ατόμων που αντιστοιχούν σε κάθε διάστημα (e_{i-1}, e_i). Για τον υπολογισμό του μέσου όρου σ' αυτή την περίπτωση αντικαθιστούμε τη συνεχή μεταβλητή με μία ασυνεχή μεταβλητή, η οποία έχει ως τιμές τα κέντρα c_i των διαστημάτων (e_{i-1}, e_i) και ως απόλυτες συχνότητες τις αντίστοιχες απόλυτες συχνότητες ν_i.

Μια πολύ βασική και χρήσιμη ιδιότητα του μέσου όρου είναι η *γραμμικότητά* του. Λέμε ότι δύο στατιστικές μεταβλητές X και X' βρίσκονται σε γραμμική σχέση μεταξύ τους (και σημειώνουμε συμβολικά

$X' = \alpha X + \beta$) αν οι τιμές της δεύτερης (X_i') συνδέονται με τιμές της πρώτης (X_i) με μια σχέση της μορφής $x_i' = \alpha x_i + \beta$, $i = 1, \ldots, \kappa$, όπου τα α και β είναι δύο σταθεροί αριθμοί, κι επίσης αν οι σχετικές συχνότητες f_i' και f_i που αντιστοιχούν στις τιμές x_i' και x_i είναι ίσες μεταξύ τους για κάθε i, δηλαδή $f_i' = f_i$, $i = 1, \ldots, \kappa$.

Η γραμμικότητα του μέσου όρου σημαίνει ότι, αν δύο στατιστικές μεταβλητές X και X' βρίσκονται σε γραμμική σχέση μεταξύ τους, οι αντίστοιχοι μέσοι όροι, $\bar{x}$ και $\bar{x}'$, συνδέονται επίσης με την ίδια σχέση· δηλαδή, διατηρώντας τους προηγούμενους συμβολισμούς, θα έχουμε:

$$\bar{x}' = \alpha\bar{x} + \beta.$$

Η γραμμικότητα του μέσου όρου αποδεικνύεται πολύ εύκολα. Πραγματικά:

$$\bar{x}' = \sum_{i=1}^{\kappa} f_i' x' = \sum_{i=1}^{\kappa} f_i(\alpha x_i + \beta) =$$

$$= \alpha\sum_{i=1}^{\kappa} f_i x_i + \beta\sum_{i=1}^{\kappa} f_i = \alpha\bar{x} + \beta,$$

μια και $\sum_{i=1}^{\kappa} f_i x_i = \bar{x}$ και $\sum_{i=1}^{\kappa} f_i = 1$.

Η γραμμικότητα του μέσου όρου χρησιμοποιείται αρκετά συχνά για τον πρακτικό υπολογισμό του, ιδιαίτερα όταν ο άμεσος υπολογισμός του με βάση τη σχέση (1) συνεπάγεται μεγάλες πράξεις. Έτσι, αντί να υπολογίσουμε τον μέσο όρο ($\bar{x}$) μιας στατιστικής μεταβλητής $\bar{x}$, υπολογίζουμε τον μέσο όρο ($\bar{x}'$) μιας άλλης μεταβλητής $\bar{x}' = \alpha x + \beta$, οπότε για τον μέσο όρο $\bar{x}$ έχουμε $\bar{x} = \frac{\bar{x}' - \alpha}{\beta}$.

Ο μέσος όρος της μεταβλητής x_i υπολογίζεται τόσο πιο εύκολα όσο

- οι τιμές x_i' είναι αριθμοί ακέραιοι,
- οι αριθμοί αυτοί είναι μικροί σε απόλυτη τιμή ώστε να είναι εύκολος ο υπολογισμός των γινομένων $f_i x_i$ ή των γινομένων $\nu_i x_i$ — αν ο υπολογισμός του μέσου όρου γίνει κατευθείαν με βάση τις απόλυτες συχνότητες.

4.1.2. Ένα άλλο μέτρο κεντρικής τάσης που χρησιμοποιείται αρκετά συχνά είναι η *διάμεσος*. Διάμεσος μιας στατιστικής μεταβλητής ονομάζεται η τιμή της μεταβλητής που χωρίζει τον πληθυσμό, διατεταγμένο ως προς τις τιμές της μεταβλητής κατά αύξουσα (ή φθίνουσα) τάξη μεγέθους, σε δύο ισοπληθή σύνολα ατόμων. Αν π.χ. έχουμε έναν πληθυσμό πέντε ατόμων και ως μεταβλητή θεωρήσουμε το ύψος τους, αφού διατάξουμε τα άτομα σύμφωνα με το ύψος τους, η διάμεσος τιμή ύψους θα είναι το ύψος του τρίτου ατόμου.

Σε αντίθεση με τον μέσο όρο, η διάμεσος ισούται πάντα με μια τιμή της μεταβλητής όχι μόνο στην περίπτωση των συνεχών μεταβλητών, αλλά και στην περίπτωση των ασυνεχών.

Όταν η στατιστική μεταβλητή είναι ασυνεχής, η διάμεσος ισούται με την τιμή της μεταβλητής αριστερά της οποίας η αθροιστική συνάρτηση είναι μικρότερη από $\frac{1}{2}$ και δεξιά της οποίας είναι μεγαλύτερη από $\frac{1}{2}$. Π.χ., η διάμεσος της στατιστικής μεταβλητής του παραδείγματος της §3.1. ισούται με 2 (βλ. και Σχ. 3.2.).

Αν ο αριθμός των ατόμων του πληθυσμού είναι άρτιος, είναι δυνατόν —αν και αρκετά σπάνιο— στην τιμή $\frac{1}{2}$ της αθροιστικής συνάρτησης ν' αντιστοιχούν δύο διαδοχικές τιμές x_i και x_{i+1} της στατιστικής μεταβλητής. Σ' αυτή την περίπτωση, όπου η διάμεσος είναι απροσδιόριστη, το διάστημα (x_i, x_{i+1}) ονομάζεται *διάμεσο διάστημα*.

Όταν η στατιστική μεταβλητή είναι συνεχής, η διάμεσος ισούται με την τιμή της στατιστικής μεταβλητής που αντιστοιχεί στην τιμή $\frac{1}{2}$ της αθροιστικής συνάρτησης. Όμως, όπως είδαμε στην §3.2., κατά κανόνα οι τιμές της αθροιστικής συνάρτησης μιας συνεχούς μεταβλητής μάς είναι γνωστές μόνο για τα άκρα $e_i (i = 0, \ldots, \kappa)$ των τάξεων των τιμών της μεταβλητής. Για τον προσδιορισμό λοιπόν της διαμέσου $\left(\text{εκτός από την ειδική περίπτωση που για κάποιο } e_\lambda \text{ έχουμε } F(e_\lambda) = \frac{1}{2}\right)$, εντοπίζουμε τις τιμές $F(e_{i-1})$ και $F(e_i)$, μεταξύ των οποίων περιέχεται ο αριθμός $\frac{1}{2}$. Η διάμεσος θα βρίσκεται επομένως μεταξύ των αντίστοιχων τιμών e_{i-1} και

e_i. Γι' αυτόν το λόγο, το διάστημα (e_{i-1}), e_i ονομάζεται *διάμεσο διάστημα της στατιστικής μεταβλητής*.

Για έναν σχετικά ακριβή υπολογισμό της διαμέσου, θεωρούμε ότι η αθροιστική συνάρτηση αυξάνει γραμμικά μεταξύ e_{i-1} και e_i, οπότε μπορούμε να προσδιορίσουμε γραφικά τη διάμεσο χρησιμοποιώντας το πολύγωνο της αθροιστικής συχνότητας (βλ. Σχ. 3.3.) ή αλγεβρικά με βάση τη σχέση:

$$\frac{M - e_{i-1}}{e_i - e_{i-1}} = \frac{1/2 - F(e_{i-1})}{F(e_i) - F(e_{i-1})},$$

απ' όπου βρίσκουμε

$$M = e_{i-1} + \alpha_i \frac{1/2 - F(e_{i-1})}{f_i}.$$

Επομένως, η διάμεσος χωρίζει το ιστόγραμμα σε δύο ίσα μέρη.

4.1.3. Μία άλλη παράμετρος κεντρικής τάσης είναι ο *τύπος* ή *σημείο μέγιστης συχνότητας* ή *επικρατέστερη τιμή* (mode). Ο τύπος είναι η τιμή της στατιστικής μεταβλητής που αντιστοιχεί στο μέγιστο του ιστογράμματος, δηλαδή η τιμή της στατιστικής μεταβλητής στην οποία αντιστοιχεί η μεγαλύτερη σχετική συχνότητα.

Όταν η στατιστική μεταβλητή είναι ασυνεχής, μπορούμε εύκολα να προσδιορίσουμε το σημείο μέγιστης συχνότητας. Π.χ., για τη μεταβλητή του παραδείγματος της §3.1., το σημείο μέγιστης συχνότητας είναι το 1. Όταν όμως η στατιστική μεταβλητή είναι συνεχής και δεν γνωρίζουμε παρά μόνο τον αριθμό των ατόμων που αντιστοιχούν σε κάθε τάξη της μεταβλητής, είναι φυσικά αδύνατο να προσδιορίσουμε με ακρίβεια το σημείο μέγιστης συχνότητας. Σ' αυτή την περίπτωση προσδιορίζουμε την τάξη μέγιστης συχνότητας η οποία ορίζεται ως η τάξη που αντιστοιχεί στη μεγαλύτερη σχετική συχνότητα ανά μονάδα πλάτους. Π.χ., για τη μεταβλητή του παραδείγματος της §3.2., η τάξη μέγιστης συχνότητας είναι η τάξη $0 - 10$ στρ., που παρουσιάζει τη μεγαλύτερη σχετική συχνότητα ανά μονάδα πλάτους, και όχι η τάξη $10 - 30$ στρ., η οποία παρουσιάζει μεν μεγαλύτερη σχετική συχνότητα, αλλά για πλάτος διπλάσιο απ' αυτό της προηγούμενης (βλ. και Σχ. 3.3.).

Τόσο για τις συνεχείς όσο και για τις ασυνεχείς στατιστικές μεταβλητές υπάρχουν περιπτώσεις όπου ο προσδιορισμός του τύπου δεν είναι μονοσήμαντος. Αυτό συμβαίνει όταν στη μεγαλύτερη σχετική συχνότητα αντιστοιχούν δύο (ή περισσότερα) σημεία ή τάξεις.

4.2. *Μέτρα Διασποράς*

4.2.1. Τα κυριότερα μέτρα διασποράς είναι η *διακύμανση* (variance) και η *τυπική απόκλιση* (standard deviation, écart type). Η διακύμανση (V) μιας στατιστικής μεταβλητής X ισούται με το σταθμισμένο (με τις σχετικές συχνότητες f_i) άθροισμα των τετραγώνων των διαφορών των τιμών της μεταβλητής από τον μέσο όρο της μεταβλητής. Δηλαδή, διατηρώντας τους συμβολισμούς που χρησιμοποιήσαμε προηγουμένως,

$$V=\sum_{i=1}^{\kappa} f_i (x_i-\bar{x})^2 \qquad (1)$$

Τη διακύμανση τη συμβολίζουμε επίσης και με $V(X)$ όταν θέλουμε να προσδιορίσουμε για ποια στατιστική μεταβλητή πρόκειται.

Επειδή η διακύμανση είναι (σταθμισμένο) άθροισμα τετραγώνων, για να εκφραζόμαστε στις μονάδες μέτρησης των τιμών της στατιστικής μεταβλητής χρησιμοποιούμε πιο συχνά, αντί για τη διακύμανση, τη θετική τετραγωνική της ρίζα, η οποία ονομάζεται *τυπική απόκλιση* και συμβολίζεται με σ:

$$\sigma=\sqrt{\sum_{i}^{\kappa} f_i (x_i-\bar{x})^2}.$$

Για τον υπολογισμό της διακύμανσης, κι επομένως και της τυπικής απόκλισης, αντί για τον τύπο (1) χρησιμοποιούμε κατά κανόνα τον τύπο:

$$V=\sum_{i=1}^{\kappa} f_i x_i^2-\bar{x}^2) \qquad (2)$$

για την εφαρμογή του οποίου δεν χρειάζεται να έχουμε προηγουμένως υπολογίσει τις διαφορές $x_i-\bar{x}$.

Ο τύπος (2) αποδεικνύεται αρκετά εύκολα. Πράγματι:

$$V=\sum_{i=1}^{\kappa} f_i (x_i-\bar{x})^2=\sum_{i=1}^{\kappa} f_i(x_i^2-2x_i\bar{x}+\bar{x}^2)=$$

$$=\sum_{i=1}^{\kappa} f_i x_i^2-2\bar{x}\sum_{i=1}^{\kappa} f_i x_i+\bar{x}^2\sum_{i=1}^{\kappa} f_i=$$

$$=\sum_{i=1}^{\kappa} f_i x_i^2-2\bar{x}\cdot\bar{x}+\bar{x}^2=\sum_{i=1}^{\kappa} f_i x_i^2-\bar{x}^2.$$

Όπως και στην περίπτωση του μέσου όρου, έτσι και για τον υπολογισμό της διακύμανσης δεν μπορούμε να εφαρμόσουμε τους τύπους (1) και (2) όταν η στατιστική μεταβλητή είναι συνεχής και δεν γνωρίζουμε παρά μόνο τον αριθμό ν_i των ατόμων που αντιστοιχούν σε κάθε διάστημα (e_{i-1}, e_i).

Για τον υπολογισμό της διακύμανσης σ' αυτή την περίπτωση αντικαθιστούμε τη συνεχή στατιστική μεταβλητή με μια ασυνεχή μεταβλητή η οποία έχει ως τιμές τα κέντρα c_i των διαστημάτων (e_{i-1}, e_i) και για απόλυτες συχνότητες τις αντίστοιχες απόλυτες συχνότητες ν_i. Η προσέγγιση αυτή έχει ως αποτέλεσμα να αυξάνει κατά κανόνα την διακύμανση, αλλά το αντίστοιχο λάθος μειώνεται όσο μικρότερο είναι το πλάτος των διαστημάτων.

Αν δύο στατιστικές μεταβλητές X και X' διαφέρουν κατά έναν σταθερό αριθμό, δηλαδή αν $X'=X+\beta$, οι αντίστοιχες διακυμάνσεις θα είναι ίσες, δηλαδή $V(X)=V(X')$, γιατί για τον υπολογισμό της διακύμανσης χρησιμοποιούμε μόνο τις διαφορές των τιμών της στατιστικής μεταβλητής από τον μέσο όρο της. Επίσης, επειδή η διακύμανση αποτελεί (σταθμισμένο) άθροισμα τετραγώνων, είναι φανερό ότι αν μια στατιστική μεταβλητή X' είναι ίση με το γινόμενο μιας στατιστικής μεταβλητής X επί έναν σταθερό αριθμό, δηλαδή αν $X'=\alpha X$, τότε η διακύμανση της πρώτης θα είναι ίση με τη διακύμανση της δεύτερης επί το τετράγωνο του αριθμού αυτού, δηλαδή $V(X')=\alpha^2 V(X)$. Από τα παραπάνω συμπεραίνουμε ότι αν δύο στατιστικές μεταβλητές X και X' βρίσκονται σε γραμμική σχέση, δηλαδή αν $X'=\alpha X+\beta$, οι διακυμάνσεις τους θα συνδέονται με τη σχέση $V(X')=\alpha^2 V(X)$, και οι αντίστοιχες τυπικές αποκλίσεις σ και σ' θα συνδέονται με τη σχέση $\sigma'=|\alpha|\sigma$. Γενικότερα, αν $X'=\sum_{i=1}^{\kappa}\alpha_i X_i+\beta$, οι αντίστοιχες διακυμάνσεις θα συνδέονται με τη σχέση $V(X')=\sum_{i=1}^{\kappa}\alpha_i^2 V(X_i)$.

Την ιδιότητα αυτή (όπως αντίστοιχα τη γραμμικότητα του μέσου όρου) χρησιμοποιούμε αρκετά συχνά για τον πρακτικό υπολογισμό της διακύμανσης και της τυπικής απόκλισης. Η διακύμανση και η τυπική

απόκλιση είναι οι κυριότερες παράμετροι διασποράς, και η χρησιμότητά τους είναι ιδιαίτερα σημαντική, γιατί προσδιορίζουν το βαθμό συγκέντρωσης των τιμών της μεταβλητής γύρω από τον μέσο όρο. Τα βασικά τους ελαττώματα είναι η σχετική πολυπλοκότητα των υπολογισμών και το γεγονός ότι είναι ευαίσθητες στις «διακυμάνσεις της δειγματοληψίας» ή στις λανθασμένες τιμές (περισσότερο από τον μέσο όρο, λόγω των τετραγώνων που εμφανίζονται στον τύπο της διακύμανσης).

4.2.2. *Συντελεστής Μεταβλητικότητας*

Ονομάζουμε *συντελεστή μεταβλητικότητας* (*CV*) μιας στατιστικής μεταβλητής το πηλίκο της τυπικής απόκλισης προς τον μέσο όρο της μεταβλητής, δηλαδή $CV = \frac{\sigma}{\bar{x}}$.

Ο συντελεστής μεταβλητικότητας είναι μια ποσότητα χωρίς διαστάσεις και ανεξάρτητη από τις μονάδες στις οποίες εκφράζεται η στατιστική μεταβλητή.

4.2.3. *Εύρος (Range, Étendue)*

Ονομάζουμε *εύρος* (*R*) τη διαφορά της μικρότερης από τη μεγαλύτερη τιμή της στατιστικής μεταβλητής, δηλαδή $R = x_{\max} - x_{\min}$.

Το εύρος είναι το απλούστερο μέτρο διασποράς, και για τον υπολογισμό του δεν χρειάζεται ούτε καν η κατάταξη όλων των τιμών της μεταβλητής. Τα κύρια μειονεκτήματά του είναι ότι βασίζεται μόνο στις δύο ακραίες τιμές, και γι' αυτό είναι ιδιαίτερα ευαίσθητο στις διακυμάνσεις της δειγματοληψίας ή στις λανθασμένες τιμές.

ΚΕΦΑΛΑΙΟ ΠΕΜΠΤΟ

ΠΕΡΙΓΡΑΦΗ ΕΝΟΣ ΠΛΗΘΥΣΜΟΥ ΩΣ ΠΡΟΣ ΔΥΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

5.1. *Πίνακες Διπλής Εισόδου*

Ως τώρα ασχοληθήκαμε με την περιγραφή ενός πληθυσμού ν ατόμων ως προς ένα χαρακτηριστικό. Ας υποθέσουμε τώρα ότι ο πληθυσμός που μελετάμε περιγράφεται ταυτόχρονα ως προς δύο χαρακτηριστικά A και B (ποιοτικά ή ποσοτικά). Συμβολίζουμε με $A_1, \ldots, A_i, \ldots, A_\kappa$ τις κ τάξεις του χαρακτηριστικού A, και με $B_1, \ldots, B_\lambda, \ldots, B_\rho$ τις ρ τάξεις του χαρακτηριστικού B. Τον αριθμό των ατόμων του πληθυσμού που ανήκουν ταυτόχρονα στην τάξη A_i του χαρακτηριστικού A και στην τάξη B_λ του χαρακτηριστικού B τον συμβολίζουμε με $\nu_{i\lambda}$, και τον ονομάζουμε *απόλυτη συχνότητα* του ζευγαριού A_i και B_λ.

Επειδή οι τάξεις του χαρακτηριστικού A, όπως κι αυτές του χαρακτηριστικού B, είναι ασυμβίβαστες μεταξύ τους και καλύπτουν όλες τις δυνατές καταστάσεις, το συνολικό άθροισμα των απόλυτων συχνοτήτων $\nu_{i\lambda}$ είναι ίσο με το μέγεθος του πληθυσμού:

$$\sum_{i=1}^{\kappa} \sum_{\lambda=1}^{\rho} \nu_{i\lambda} = \nu \qquad (1)$$

Ονομάζουμε *σχετική συχνότητα* του ζευγαριού των τάξεων A_i και B_λ, και τη συμβολίζουμε με $f_{i\lambda}$, το πηλίκο της αντίστοιχης απόλυτης συχνότητας προς το συνολικό μέγεθος του πληθυσμού:

$$f_{i\lambda} = \frac{\nu_{i\lambda}}{\nu}.$$

Από τη σχέση (1) προκύπτει ότι το άθροισμα των σχετικών συχνοτήτων είναι ίσο με τη μονάδα:

$$\sum_{i=1}^{\kappa}\sum_{\lambda=1}^{\rho} f_{i\lambda}=1.$$

Ο στατιστικός πίνακας που περιγράφει τα ν άτομα του πληθυσμού ταυτόχρονα και ως προς τα δύο χαρακτηριστικά ονομάζεται *πίνακας διπλής εισόδου*. Οι γραμμές αντιστοιχούν στις τάξεις του χαρακτηριστικού A και οι στήλες στις τάξεις του χαρακτηριστικού B. Στην τομή της γραμμής i και της στήλης λ γράφουμε την απόλυτη συχνότητα $\nu_{i\lambda}$ του ζευγαριού A_i και B_λ.

Η γενική μορφή ενός πίνακα διπλής εισόδου είναι η εξής:

ΠΙΝΑΚΑΣ 5.1.

Τάξεις του χαρακτηριστικού B / *Τάξεις του χαρακτηριστικού A*	B_1	B_2	...	B_λ	...	B_ρ	*Σύνολο*
A_1	V_{11}	V_{12}	...	$V_{1\lambda}$	...	$V_{1\rho}$	$V_{1\cdot}$
A_2	V_{21}	V_{22}	...	$V_{2\lambda}$	...	$V_{2\rho}$	$V_{2\cdot}$
⋮	..	..	...	...	...	...	
A_i	V_{i1}	V_{i2}	...	$V_{i\lambda}$	...	$V_{i\rho}$	$V_{i\cdot}$
⋮	...	...	..	...	...	...	
A_κ	$V_{\kappa 1}$	$V_{\kappa 2}$	...	$V_{\kappa\lambda}$	...	$V_{\kappa\rho}$	$V_{\kappa\cdot}$
Σύνολο	$V_{\cdot 1}$	$V_{\cdot 1}$	...	$V_{\cdot\lambda}$	...	$V_{\cdot\rho}$	V

Με μία τελεία στη θέση του αντίστοιχου δείκτη συμβολίζουμε το μερικό άθροισμα ως προς τον δείκτη αυτό: $\nu_{i\cdot}$. Είναι το άθροισμα των $\nu_{i\lambda}$ για τα οποία το i είναι σταθερό, δηλαδή ο αριθμός των ατόμων του πληθυσμού που ανήκουν στην τάξη A_i, ανεξάρτητα από το σε ποια τάξη του χαρακτηριστικού B ανήκουν. Αντίστοιχα ορίζουμε και τα $\nu_{\cdot\lambda}$.

Έχουμε λοιπόν:

$$\nu_{i\cdot}=\sum_{\lambda=1}^{\rho}\nu_{i\lambda} \text{ και } \nu_{\cdot\lambda}=\sum_{i=1}^{\kappa}\nu_{i\lambda},$$

κι επομένως:

$$\sum_{i=1}^{\kappa}\nu_{i\cdot}=\sum_{\lambda=1}^{\rho}\nu_{\cdot\lambda}=\nu.$$

Τα $\nu_{i\cdot}$ για τους διάφορους δείκτες i ($i=1, \ldots, \kappa$) αποτελούν τις απόλυτες συχνότητες της λεγόμενης *περιθωριακής* (marginale) *κατανομής* ως προς το χαρακτηριστικό A. Αντίστοιχα, τα $\nu_{\cdot\lambda}$ ($\lambda=1, \ldots, \rho$) ορίζουν την περιθωριακή κατανομή ως προς το χαρακτηριστικό B. Οι σχετικές συχνότητες των κατανομών αυτών, $f_{i\cdot}$ ($i=1, \ldots, \kappa$) και $f_{\cdot\lambda}$ ($\lambda=1, \ldots, \rho$) αντίστοιχα, ορίζονται επομένως από τις σχέσεις:

$$f_{i\cdot}=\frac{\nu_{i\cdot}}{\nu}=\sum_{\lambda=1}^{\rho} f_{i\lambda} \qquad \text{και} \qquad f_{\cdot\lambda}=\frac{\nu_{\cdot\lambda}}{\nu}=\sum_{i=1}^{\kappa} f_{i\lambda}.$$

Εκτός από τις δύο περιθωριακές κατανομές, μπορούμε, με βάση τον πίνακα διπλής εισόδου, να ορίσουμε και μια σειρά άλλες κατανομές που παρουσιάζουν ιδιαίτερο ενδιαφέρον. Πραγματικά, κάθε γραμμή και κάθε στήλη του πίνακα διπλής εισόδου περιγράφει έναν συγκεκριμένο υποπληθυσμό: ή λ στήλη π.χ. περιγράφει, ως προς τις τάξεις του χαρακτηριστικού A, τον υποπληθυσμό που ανήκει στην τάξη B_λ· το μέγεθος του υποπληθυσμού αυτού είναι προφανώς $\nu_{\cdot\lambda}$. Αν τώρα περιορίσουμε τη μελέτη μας σ'αυτόν τον υποπληθυσμό (δηλαδή στα άτομα που ανήκουν στην τάξη B_λ), τα $\nu_{i\lambda}$ εκφράζουν τις απόλυτες συχνότητες των τάξεων A_i ($=i$, $\ldots, \kappa$), κι επομένως τα πηλίκα $f_i^\lambda=\frac{\nu_{i\lambda}}{\nu^\lambda}$ ($i=1, \ldots, \kappa$) των απόλυτων συχνοτήτων προς το μέγεθος του υποπληθυσμού ορίζουν τις αντίστοιχες σχετικές συχνότητες, τις οποίες ονομάζουμε *υπό συνθήκη σχετικές συχνότητες* των τάξεων A_i ($i=1, \ldots, \kappa$) ως προς την τάξη B_λ.

Η κατανομή που ορίζεται μ' αυτό τον τρόπο ονομάζεται *υπό συνθήκη κατανομή* ως προς το χαρακτηριστικό A των ατόμων που ανήκουν στην τάξη B_λ του χαρακτηριστικού B. Υπάρχουν προφανώς ρ υπό συνθήκη κατανομές ως προς το χαρακτηριστικό A, που η καθεμιά τους αντιστοιχεί σε μια τάξη B_λ ($\lambda=1, \ldots, \rho$) του χαρακτηριστικού B.

Όπως ορίζονται οι υπό συνθήκη σχετικές συχνότητες των τάξεων A_i ως προς μία τάξη B_λ, αντίστοιχα ορίζονται και οι υπό συνθήκη σχετικές συχνότητες (f_λ^i) των τάξεων B_λ ως προς μία τάξη A_i, κι έχουμε $f_\lambda^i = \frac{\nu_{i\lambda}}{\nu_i}$.

Υπάρχουν προφανώς κ υπό συνθήκη κατανομές ως προς το χαρακτηριστικό B, που η καθεμιά τους αντιστοιχεί σε μια τάξη A_i ($i = 1, \ldots, \kappa$) του χαρακτηριστικού A.

Η σχετική συχνότητα $f_{i\lambda}$ της συνολικής κατανομής ως προς τα δύο χαρακτηριστικά συνδέεται με τις αντίστοιχες περιθωριακές σχετικές συχνότητες (f_i και f_λ) και υπό συνθήκη σχετικές συχνότητες (f_λ^i και f_i^λ) με τις εξής σχέσεις:

$$f_{i\lambda} = \frac{\nu_{i\lambda}}{\nu} = \frac{\nu_{i.}}{\nu} \cdot \frac{\nu_{i\lambda}}{\nu_{i.}} = f_{i.} \cdot f_\lambda^i \qquad (1)$$

$$f_{i\lambda} = \frac{\nu_\lambda}{\nu} = \frac{\nu_{.\lambda}}{\nu} \cdot \frac{\nu_{i\lambda}}{\nu_{.\lambda}} = f_{.\lambda} \cdot f_i^\lambda \qquad (2)$$

Παραδείγματος χάρη:

ΠΙΝΑΚΑΣ 5.2.

ΚΑΤΑΝΟΜΗ ΤΟΥ ΠΛΗΘΥΣΜΟΥ ΤΗΣ ΕΛΛΑΔΑΣ ΩΣ ΠΡΟΣ ΤΗ ΘΡΗΣΚΕΙΑ ΚΑΙ ΤΗ ΓΛΩΣΣΑ ΤΟ 1968 (ΣΤΟΙΧΕΙΑ ΤΗΣ ΑΠΟΓΡΑΦΗΣ ΤΟΥ 1928).

ΘΡΗΣΚΕΥΜΑ / ΓΛΩΣΣΑ	Σύνολο	Χριστιανοί			Μουσουλμάνοι	Ισραηλίται	Λοιπών θρησκειών	Ουδεμιάς θρησκείας
		Ορθόδοξοι	Καθολικοί	Διαμαρτυρόμενοι				
		Απόλυτοι αριθμοί						
Ελληνική	5.759.523	5.716.100	27.747	3.867	2.623	9.090	15	81
Τουρκική	191.254	103.642	327	760	86.506	17	1	1
Μακεδονοσλαβική	81.984	81.844	68	11	2	58	—	1
Ισπανική	63.200	28	58	41	72	62.999	—	2
Αρμενική	33.631	31.038	1.432	16	10	10	2	—
Κουτσοβλαχική	19.703	19.679	9	2	3	10	—	—
Αλβανική	18.773	95	59	17	18.598	3	1	—

ΘΡΗΣΚΕΥΜΑ / ΓΛΩΣΣΑ	Σύνολο	Χριστιανοί			Μουσουλμάνοι	Ισραηλίται	Λοιπών θρησκειών	Ουδεμιάς θρησκείας
		Ορθόδοξοι	Καθολικοί	Διαμαρτυρόμενοι				
		Απόλυτοι αριθμοί						
Βουλγαρική	16.775	20	—	—	16.755	—	—	—
Αθιγγανική	4.998	3.853	—	1	1.130	—	14	—
Ρωσική	3.295	3.177	49	14	3	40	—	12
Ιταλική	3.199	98	2.878	18	1	203	—	1
Αγγλική	2.098	201	274	1.605	1	15	—	2
Λοιπαί ξέναι γλώσσαι	6.248	1.751	2.577	1.235	317	316	12	17
Σύνολο	6.204.634	5.961.529	35.182	9.003	123.017	72.791	45	117

ΠΗΓΗ: Στατιστική Επετηρίς της Ελλάδος (1930, σ. 98).

Οι σχετικές συχνότητες (σε ‰) της περιθωριακής κατανομής ως προς το χαρακτηριστικό A (γλώσσα) εκφράζουν την αναλογία κάθε γλώσσας επί 1.000 ατόμων του συνολικού πληθυσμού, και οι σχετικές συχνότητες (σε ‰) καθεμιάς από τις υπό συνθήκη κατανομές ως προς το χαρακτηριστικό A την αναλογία κάθε γλώσσας επί 1.000 ατόμων του συγκεκριμένου θρησκεύματος. Ο αντίστοιχος πίνακας είναι ο εξής:

ΠΙΝΑΚΑΣ 5.3.

ΘΡΗΣΚΕΥΜΑ / ΓΛΩΣΣΑ	Σύνολο	Χριστιανοί			Μουσουλμάνοι	Ισραηλίται	Λοιπών θρησκειών	Ουδεμιάς θρησκείας
		Ορθόδοξοι	Καθολικοί	Διαμαρτυρόμενοι				
		Αναλογία κάθε γλώσσας επί 1.000 ατόμων						
Ελληνική	928.25	958.83	788.67	429.52	20.81	124.88	333.33	692.31
Τουρκική	30.83	17.39	9.29	84.42	686.46	0.23	22.22	8.55
Μακεδονοσλαβική	13.21	13.73	1.93	1.22	0.02	0.80	—	8.55
Ισπανική	10.19	—	1.65	4.56	0.57	865.48	—	17.09
Αρμενική	5.42	5.21	32.29	159.06	0.13	0.13	44.45	—
Κουτσοβλαχική	3.18	3.30	0.26	0.22	0.02	0.14	—	—
Αλβανική	3.02	0.02	1.68	1.89	147.58	0.04	22.22	—

ΘΡΗΣΚΕΥΜΑ / ΓΛΩΣΣΑ	Σύνολο	Χριστιανοί			Μουσουλμάνοι	Ισραηλίται	Λοιπών θρησκειών	Ουδεμιάς θρησκείας
		Ορθόδοξοι	Καθολικοί	Διαμαρτυρόμενοι				
		Αναλογία κάθε γλώσσας επί 1.000 ατόμων						
Βουλγαρική	2.70	—	—	—	132.96	—	—	—
Αθιγγανική	0.81	0.65	—	0.11	8.97	—	311.11	—
Ρωσική	0.53	0.53	1.39	1.55	0.02	0.55	—	102.56
Ιταλική	0.51	0.02	81.80	2.00	0.01	2.79	—	8.55
Αγγλική	0.34	0.03	7.79	178.27	0.01	0.21	—	17.09
Λοιπαί ξέναι γλώσσαι	1.01	0.29	73.25	137.18	2.44	4.75	266.67	145.30
Σύνολο	1000.00	1000.00	1000.00	1000.00	1000.000	1000.00	1000.00	1000.00

ΠΗΓΗ: ό.π.

Οι σχετικές συχνότητες (σε ‰) της περιθωριακής κατανομής ως προς το χαρακτηριστικό *B* (θρησκεία) εκφράζουν την αναλογία κάθε θρησκείας επί 1.000 ατόμων του συνολικού πληθυσμού, και οι σχετικές συχνότητες (σε ‰) καθεμιάς από τις υπό συνθήκη κατανομές ως προς το χαρακτηριστικό *B* την αναλογία κάθε θρησκείας επί 1.000 ατόμων της συγκεκριμένης γλώσσας. Ο αντίστοιχος πίνακας είναι ο εξής:

ΠΙΝΑΚΑΣ 5.4.

ΘΡΗΣΚΕΥΜΑ / ΓΛΩΣΣΑ	Σύνολο	Χριστιανοί			Μουσουλμάνοι	Ισραηλίται	Λοιπών θρησκειών	Ουδεμιάς θρησκείας
		Ορθόδοξοι	Καθολικοί	Διαμαρτυρόμενοι				
		Αναλογία κάθε θρησκείας επί 1.000 ατόμων						
Ελληνική	1000.00	992.46	4.82	0.67	0.46	1.58	—	0.01
Τουρκική	1000.00	541.91	1.71	3.98	452.31	0.09	—	—
Μακεδονοσλαβική	1000.00	998.29	0.83	0.13	0.03	0.71	—	0.01
Ισπανική	1000.00	0.44	0.92	0.65	1.14	996.82	—	0.03
Αρμενική	100.000	992.82	33.77	42.57	0.48	0.30	0.06	—
Κουτσοβλαχική	1000.00	998.78	0.46	0.10	0.15	0.51	—	—
Αλβανική	1000.00	5.06	3.14	0.91	990.68	0.16	0.05	—
Βουλγαρική	1000.00	1.19	—	—	998.81	—	—	—
Αθιγγανική	1000.00	770.91	—	0.20	226.09	—	2.80	—

ΘΡΗΣΚΕΥΜΑ / ΓΛΩΣΣΑ	Σύνολο	Χριστιανοί			Μουσουλμάνοι	Ισραηλίται	Λοιπών θρησκειών	Ουδεμιάς θρησκείας
		Ορθόδοξοι	Καθολικοί	Διαμαρτυρόμενοι				
		Αναλογία κάθε θρησκείας επί 1.000 ατόμων						
Ρωσική	1000.00	964.19	14.87	4.25	0.91	12.14	—	3.64
Ιταλική	1000.00	30.63	899.66	5.63	0.31	63.46	—	0.31
Αγγλική	1000.00	95.81	130.60	765.01	0.48	7.15	—	0.95
Λοιπαί ξέναι γλώσσαι	1000.00	280.73	412.45	197.66	49.14	55.38	1.92	2.72
Σύνολο	1000.00	960.81	5.67	1.45	20.31	11.73	0.01	0.02

ΠΗΓΗ: ό.π.

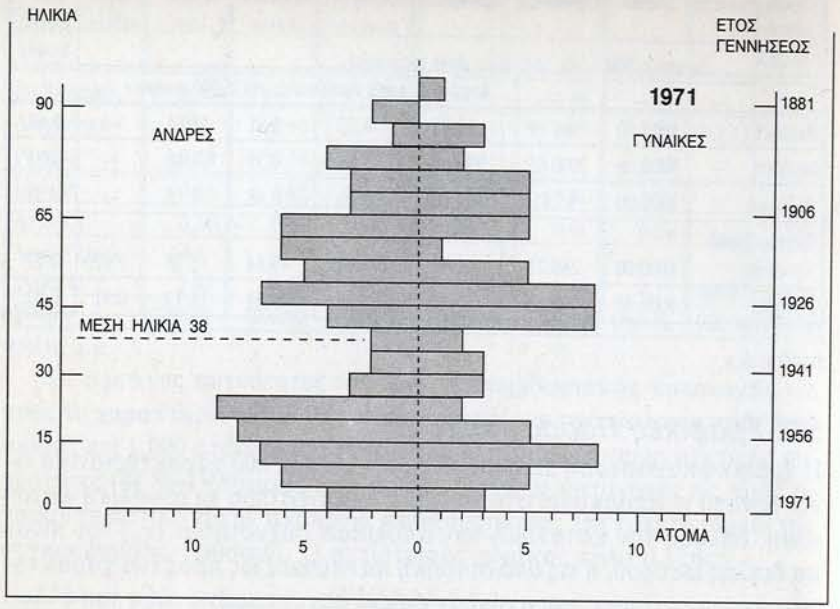
5.2. *Γραφικές Παραστάσεις*

Η γραφική παράσταση μιας κατανομής ως προς δύο χαρακτηριστικά είναι δυνατό ν' αποσκοπεί στο να κάνει εμφανέστερη τη *συνολική κατανομή*, δηλαδή την κατανομή των απόλυτων συχνοτήτων ($\nu_{i\lambda}$) του πίνακα διπλής εισόδου, ή τις *υπό συνθήκη κατανομές* ως προς ένα χαρακτηριστικό των ατόμων που ανήκουν σε μια συγκεκριμένη τάξη του άλλου χαρακτηριστικού. Επιπλέον, η γραφική παράσταση μιας κατανομής ως προς δύο χαρακτηριστικά συχνά διαφέρει ανάλογα με τη φύση καθενός από τα δύο χαρακτηριστικά (ποιοτικά, ασυνεχή ποσοτικά, συνεχή ποσοτικά). Γι' αυτούς τους λόγους υπάρχει μια μεγάλη ποικιλία γραφικών παραστάσεων για τις κατανομές ως προς δύο χαρακτηριστικά. Στα πλαίσια του παρόντος βιβλίου δεν είναι φυσικά δυνατό να γίνει μια αναλυτική παρουσίασή τους. Γι' αυτό θα περιοριστούμε μόνο σε μερικά παραδείγματα.

5.2.1. *Πυραμίδα Ηλικιών*

Η πυραμίδα ηλικιών χρησιμοποιείται συνήθως για τη γραφική παράσταση της συνολικής κατανομής ενός πληθυσμού ως προς το φύλο και την ηλικία (βλ. Σχ. 5.1.):

ΣΧΗΜΑ 5.1. *Πυραμίδα ηλικιών των κατοίκων της νήσου Δονούσα το 1971.*

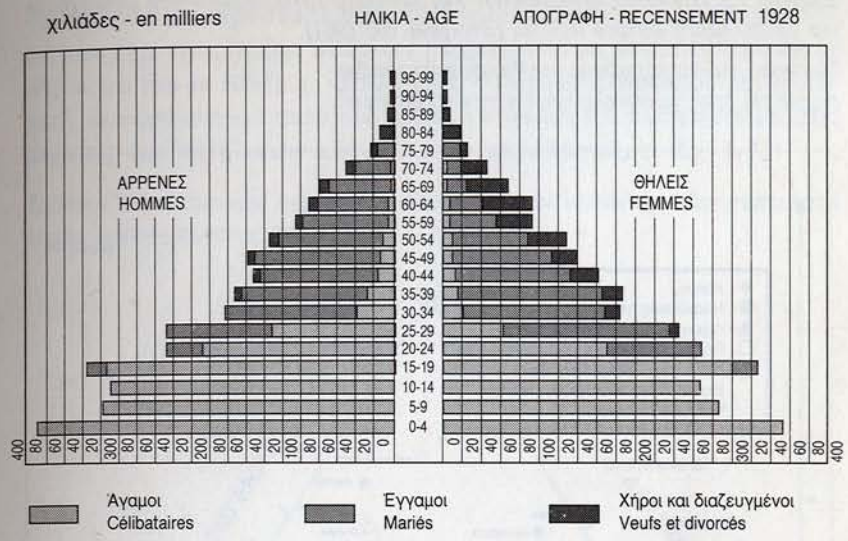


ΠΗΓΗ: E. KOLODNY, *ό.π.* (τ. 3, Διάγρ. Κ.18).

Με την πυραμίδα ηλικιών μπορούμε επίσης να παραστήσουμε τη συνολική κατανομή ενός πληθυσμού ως προς τρία χαρακτηριστικά (φύλο, ηλικία και οικογενειακή κατάσταση· βλ., για παράδειγμα, το Σχ. 5.2.).

Στο Σχ. 5.2. το αριστερό μισό της πυραμίδας, που αφορά τον ανδρικό πληθυσμό, αποτελεί τη γραφική παράσταση μιας κατανομής ως προς δύο χαρακτηριστικά (την ηλικία και την οικογενειακή κατάσταση). Ο τρόπος αυτός γραφικής παράστασης, που βασίζεται στα ακιδωτά διαγράμματα (βλ. §2.3.), είναι ο πιο συνηθισμένος στις περιπτώσεις που το ένα τουλάχιστον από τα χαρακτηριστικά είναι ποιοτικό.

ΣΧΗΜΑ 5.2. *Κατανομή του πληθυσμού της Ελλάδας ως προς την ηλικία, το φύλο και την οικογενειακή κατάσταση.*



ΠΗΓΗ: Στατιστική Επετηρίς της Ελλάδος (1930, Πίν. IV).

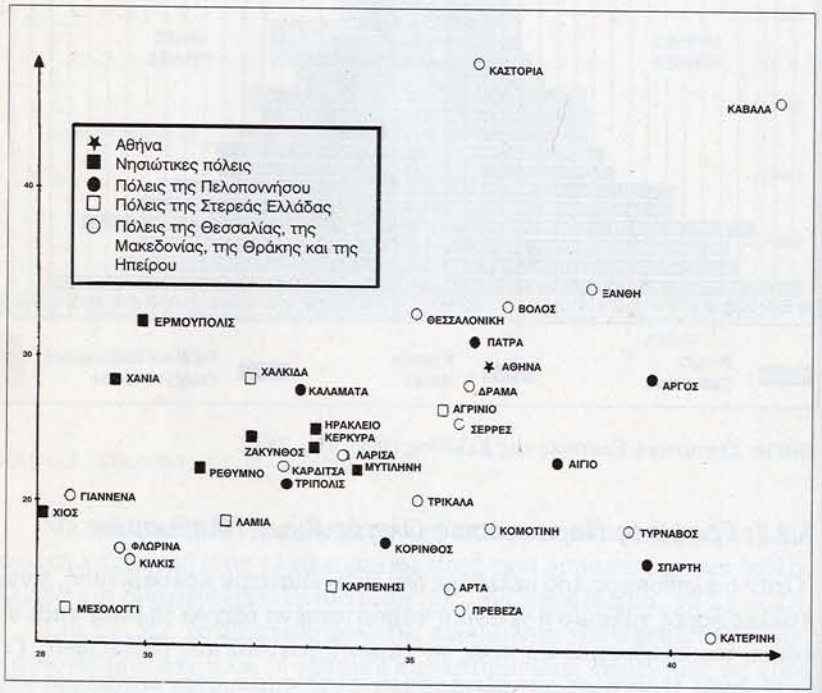
5.2.2. *Γραφικές Παραστάσεις Ολιγάριθμων Πληθυσμών*

Όταν ο πληθυσμός που μελετάμε δεν είναι ιδιαίτερα πολυάριθμος, είναι πολλές φορές χρήσιμο η γραφική παράσταση να δείχνει τη θέση κάθε ατόμου του πληθυσμού ως προς τα χαρακτηριστικά που εξετάζουμε. Γι' αυτό, στην περίπτωση των ολιγάριθμων πληθυσμών, όταν έχουμε μια κατανομή ως προς δύο ποσοτικά χαρακτηριστικά χρησιμοποιούμε συχνά ένα *καρτεσιανό σύστημα αξόνων*, όπου κάθε άτομο του πληθυσμού παριστάνεται από ένα σημείο με συντεταγμένες τις τιμές των αντίστοιχων στατιστικών μεταβλητών. Η γραφική παράσταση μιας κατανομής ως προς δύο ποσοτικά χαρακτηριστικά σε καρτεσιανό σύστημα αξόνων είναι ιδιαίτερα χρήσιμη, γιατί αποτελεί έναν άμεσο τρόπο για να εντοπί-

σουμε αν υπάρχει σχέση, και τι είδους, ανάμεσα στα δύο ποσοτικά χαρακτηριστικά (βλ.,για παράδειγμα, το Σχ. 5.3.):

ΣΧΗΜΑ 5.3. *Ποσοστό εργαζομένων και ποσοστό εργαζομένων στη βιομηχανία για τα ελληνικά αστικά κέντρα (στοιχεία του 1961).*

Ποσοστό των εργαζομένων στη βιομηχανία και βιοτεχνία ως προς το σύνολο του ενεργού πληθυσμού.

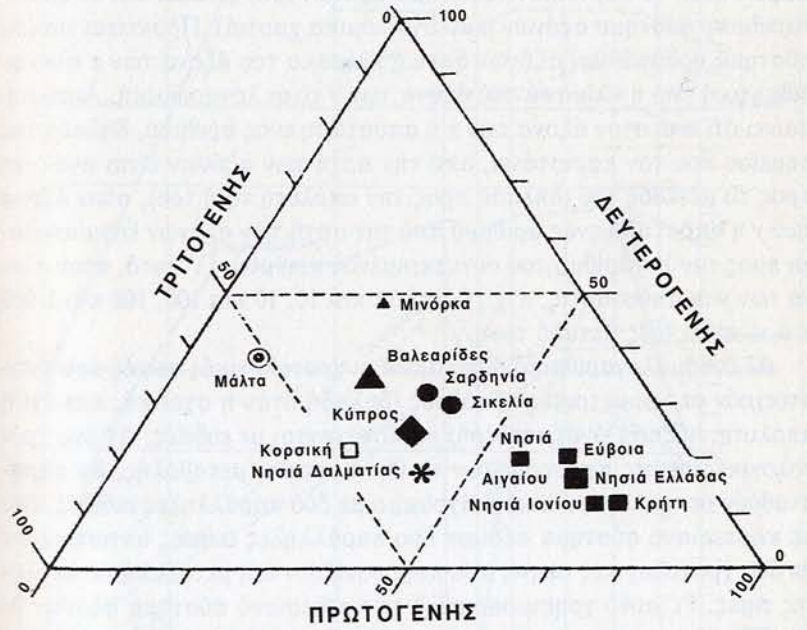


Ποσοστό απασχόλησης (στους εργαζόμενους περιλαμβάνονται και οι άνεργοι καθώς και οι νέοι που ζητούν εργασία για πρώτη φορά).

ΠΗΓΗ: GUY BURGEL, *Athènes, étude de la croissance d'une capitale méditerranéenne* (Παρίσι, 1975, Διάγρ. II.8).

Ένας άλλος τρόπος που χρησιμοποιείται συχνά στις κοινωνικές επιστήμες για τη γραφική παράσταση των ατόμων ενός πληθυσμού είναι το *τριγωνικό γράφημα*. Η πιο τυπική περίπτωση χρησιμοποίησής του είναι για τη γραφική παράσταση ενός πληθυσμού ως προς ένα χαρακτηριστικό που έχει τρεις τάξεις οι οποίες εκφράζονται με ποσοστά που έχουν άθροισμα ίσο με 100 (π.χ. κατανομή του ενεργού πληθυσμού σε πρωτογενή, δευτερογενή και τριτογενή τομέα, κατανομή του πληθυσμού σε τρεις βασικές ομάδες ηλικιών κλπ.· βλ., για παράδειγμα, το Σχ. 5.4.):

ΣΧΗΜΑ 5.4. *Κατανομή του ενεργού πληθυσμού των νήσων της Μεσογείου κατά τομέα δραστηριότητας (1957/1965).*



ΠΗΓΗ: E. KOLODNY, *ό.π.* (τ. 3, Διάγρ. G.1).

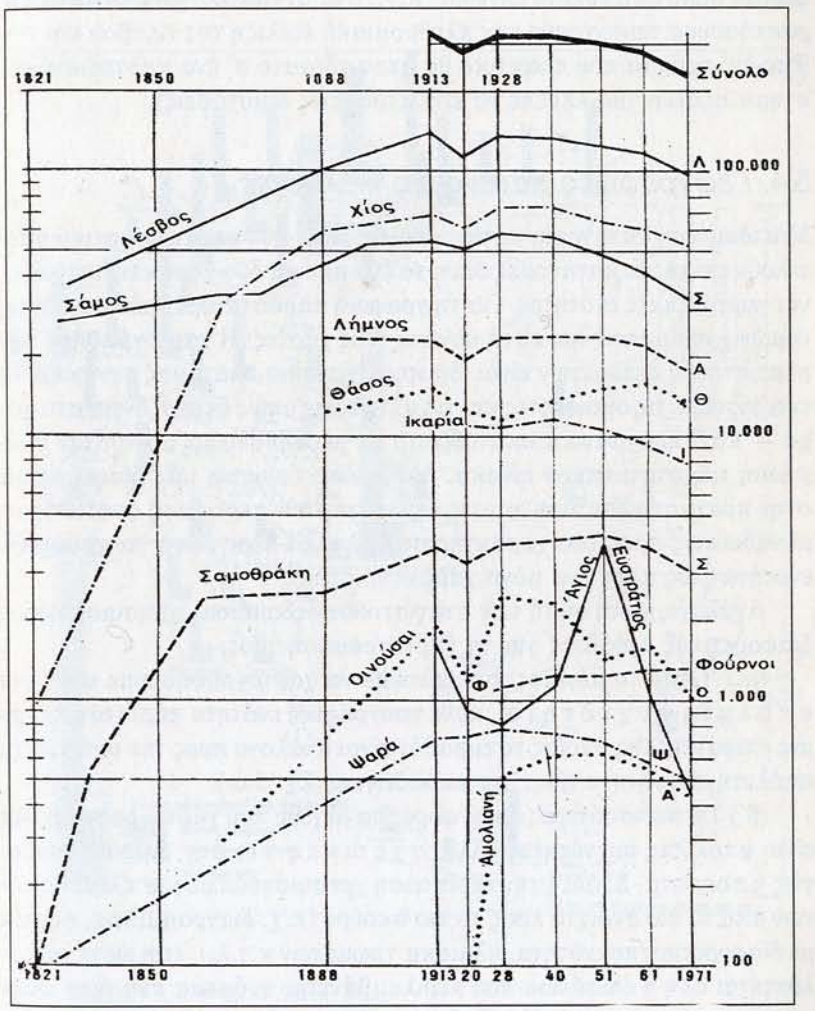
5.3. *Χρονολογικές Σειρές*

Μια ιδιαίτερη κατηγορία κατανομών ως προς δύο χαρακτηριστικά αποτελούν οι κατανομές όπου το ένα από τα δύο χαρακτηριστικά είναι ο χ ρ ό ν ο ς. Οι κατανομές αυτές ονομάζονται συνήθως *χρονολογικές σειρές*. Το δεύτερο χαρακτηριστικό μπορεί να είναι ποιοτικό ή ποσοτικό. Οι χρονολογικές σειρές χρησιμοποιούνται ευρύτατα στην Οικονομία, και γι' αυτό η σχετική θεωρία, καθώς και οι τεχνικές που εφαρμόζονται για την επεξεργασία και την παρουσίαση των δεδομένων, είναι ιδιαίτερα ανεπτυγμένες.

Οι πιο απλές μέθοδοι για τη γραφική παράσταση των χρονολογικών σειρών είναι το *καρτεσιανό σύστημα αξόνων* (βλ. §5.2.2.) και το *ημιλογαριθμικό σύστημα αξόνων* (ημιλογαριθμικά χαρτιά). Πρόκειται για ένα σύστημα ορθογώνιων αξόνων όπου η κλίμακα του άξονα των x είναι αριθμητική ενώ η κλίμακα του άξονα των y είναι λογαριθμική. Αυτό σημαίνει ότι ενώ στον άξονα των x η απόσταση ενός αριθμού, δηλαδή του σημείου που τον παριστάνει, από την αρχή των αξόνων είναι ανάλογη προς το μέγεθός του (δηλαδή προς την απόλυτη τιμή του), στον άξονα των y η απόσταση ενός αριθμού από την αρχή των αξόνων είναι ανάλογη προς τον λογάριθμο του συγκεκριμένου αριθμού. Γι' αυτό, στον άξονα των y οι αποστάσεις, π.χ., μεταξύ 1 και 10, 10 και 100, 100 και 1.000 κ.ο.κ. είναι ίσες μεταξύ τους.

Σ' ένα ημιλογαριθμικό διάγραμμα οι χρονολογικές σειρές που αντιστοιχούν σε γεωμετρικές προόδους (δηλαδή όταν η σχετική, και όχι η απόλυτη, αύξηση είναι σταθερή) παριστάνονται με ευθείες. Αν δύο χρονολογικές σειρές παρουσιάζουν το ίδιο ποσοστό μεταβολής, θα παρασταθούν σε ημιλογαριθμικό διάγραμμα με δύο παράλληλες ευθείες, ενώ σε καρτεσιανό σύστημα αξόνων δύο παράλληλες ευθείες αντιστοιχούν σε δύο χρονολογικές σειρές που παρουσιάζουν ίση μεταβολή σε απόλυτες τιμές. Γι' αυτό χρησιμοποιούμε το καρτεσιανό σύστημα αξόνων όταν θέλουμε να τονίσουμε το απόλυτο μέγεθος της μεταβολής από τη μια χρονική στιγμή στην άλλη, ενώ χρησιμοποιούμε το ημιλογαριθμικό σύστημα αξόνων όταν θέλουμε να τονίσουμε το αντίστοιχο ποσοστό μεταβολής. Το ημιλογαριθμικό σύστημα το χρησιμοποιούμε επίσης κι όταν θέ-

ΣΧΗΜΑ 5.5. *Πληθυσμιακή εξέλιξη των νήσων του Ανατολικού Αιγαίου 1821-1971.*



ΠΗΓΗ: KOLODNY, *ό.π.* (τ. 3, Διάγρ. Η.11).

λουμε να παραστήσουμε ταυτόχρονα δύο χρονολογικές σειρές που διαφέρουν πολύ σε απόλυτο μέγεθος· π.χ., στο Σχήμα 5.5. μπορούμε να παραστήσουμε ταυτόχρονα την πληθυσμιακή εξέλιξη της Λέσβου και των Ψαρών, πράγμα που πρακτικά θα ήταν αδύνατο σ' ένα καρτεσιανό σύστημα αξόνων (θα έπρεπε να είχε τεράστιες διαστάσεις).

5.4. *Γεωγραφικές Κατανομές — Χάρτες*

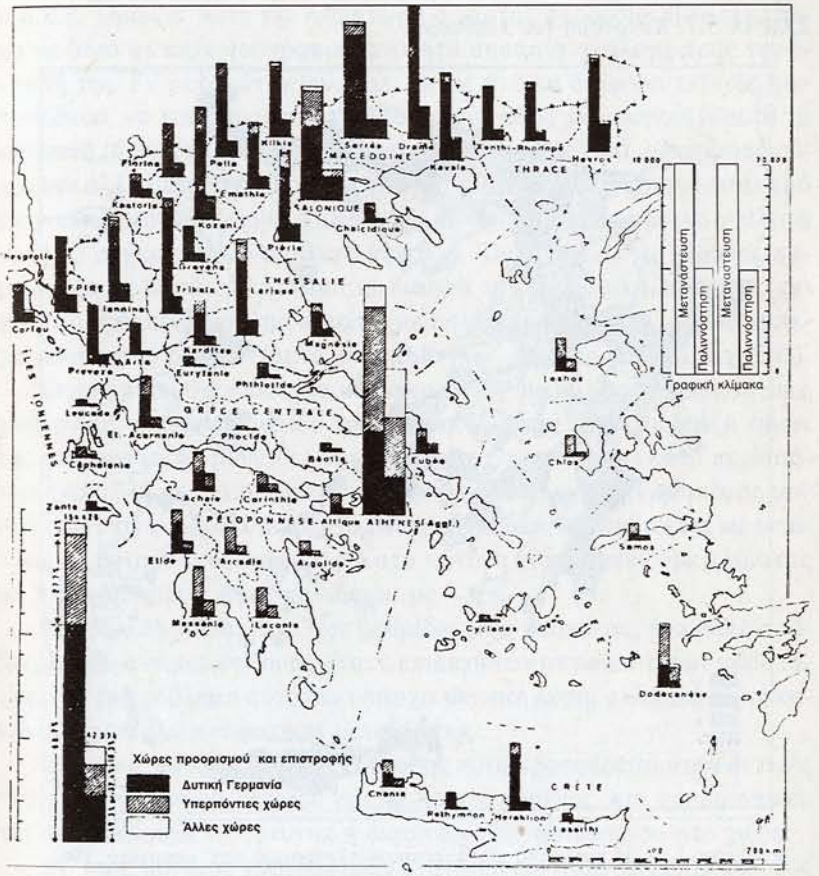
Μια ιδιαίτερη κατηγορία κατανομών ως προς δύο χαρακτηριστικά αποτελούν επίσης οι κατανομές όπου το ένα από τα δύο χαρακτηριστικά είναι γεωγραφικές ενότητες. Για τη γραφική παράσταση αυτών των κατανομών χρησιμοποιούμε κατά κανόνα τους χάρτες. Η χαρτογράφηση των στατιστικών δεδομένων είναι ιδιαίτερα χρήσιμη, γιατί μας επιτρέπει να εντοπίσουμε τις ομοιότητες και τις αντιθέσεις όπως εκφράζονται στο χώρο — κάτι πραγματικά ακατόθρωτο αν περιοριστούμε μόνο στην ανάγνωση του στατιστικού πίνακα. Δεν πρέπει όμως να μας διαφεύγει ότι στην πραγματικότητα οι χάρτες δεν αποτελούν περιγραφή στατιστικών μονάδων ως προς δύο χαρακτηριστικά, αλλά περιγραφή γεωγραφικών ενοτήτων ως προς ένα μόνο χαρακτηριστικό.

Ανάλογα με τη φύση των στατιστικών δεδομένων, χρησιμοποιούμε διαφορετικές μεθόδους για τη χαρτογράφησή τους:

(α΄) Όταν το μέγεθος που θέλουμε να χαρτογραφήσουμε είναι μια απόλυτη συχνότητα, κάθε γεωγραφική ενότητα παριστάνεται με μια επιφάνεια της οποίας το εμβαδόν είναι ανάλογο προς την αντίστοιχη απόλυτη συχνότητα (βλ., για παράδειγμα, Σχ. 5.6.).

(β΄) Τις περισσότερες όμως φορές τα μεγέθη που χαρτογραφούμε δεν είναι απόλυτες συχνότητες αλλά σχετικές τιμές, δηλαδή αναλογίες ή ποσοστά. Σ' αυτή την περίπτωση χρησιμοποιούμε μια κλίμακα τόνων από το πιο ανοιχτό προς το πιο σκούρο (π.χ. διαγραμμίσεις, σημεία με διαφορετική πυκνότητα, κλίμακα χρωμάτων κ.τ.λ.), έτσι ώστε να καλύπτεται όλη η επιφάνεια που περιλαμβάνεται ανάμεσα στα όρια κάθε γεωγραφικής ενότητας. Η μέθοδος αυτή προσφέρεται ιδιαίτερα όταν θέλουμε να χαρτογραφήσουμε πυκνότητες ανά μονάδα επιφάνειας (π.χ. πυκνότητα κατοίκησης, ποσοστό της συνολικής επιφάνειας που αφιερώ-

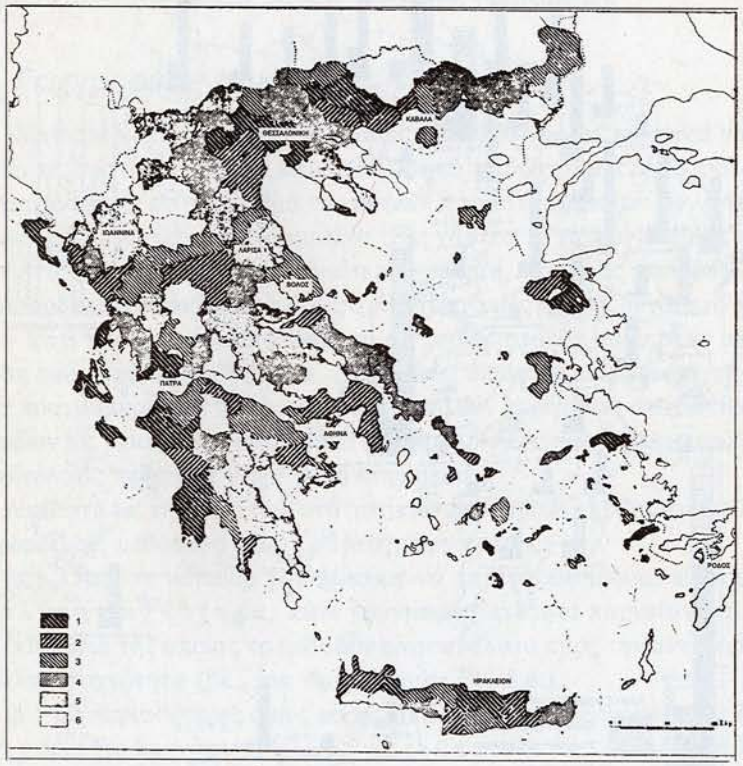
ΣΧΗΜΑ 5.6. *Μετανάστευση και παλινόστηση 1970-1971 (ανά νομό και χώρα προορισμού ή επιστροφής).*



ΠΗΓΗ: KOLODNY, *ό.π.* (τ. 3, Διάγρ. J.22).

νεται σε μια συγκεκριμένη καλλιέργεια, κ.ο.κ.· βλ., για παράδειγμα, Σχ. 5.7.):

ΣΧΗΜΑ 5.7. *Κατανομή του πληθυσμού.*



Χάρτης 3. — Η πυκνότης του αγροτικού πληθυσμού κατ' επαρχίαν, 1961.
1. 119,50 - 56,17 κάτ. ανά τετρ. χιλ. 4. 37,09 - 31,07 κάτ. ανά τετρ. χιλ.
2. 54,87 - 45,01 κάτ. ανά τετρ. χιλ. 5. 30,69 - 26,31 κάτ. ανά τετρ. χιλ.
3. 44,53 - 37,81 κάτ. ανά τετρ. χιλ. 6. 26,24 - 6,48 κάτ. ανά τετρ. χιλ.
(Χάρτης ληφθείς εκ του *Οικονομικού και Κοινωνικού Άτλαντος της Ελλάδος*).

ΠΗΓΗ: B. KAYSER, *Ανθρωπογεωγραφία της Ελλάδας* (εκδ. ΕΚΚΕ, Αθήνα, 1968, σ. 17).

Την ίδια μέθοδο χρησιμοποιούμε κατά κανόνα κι όταν το ποσοστό που θέλουμε να χαρτογραφήσουμε δεν αναφέρεται σε μονάδες επιφανείας (π.χ. ποσοστό ενεργού πληθυσμού, ποσοστό παιδικής θνησιμότητας κ.ο.κ.). Όμως σ' αυτή την περίπτωση ο χάρτης παρουσιάζει το ελάττωμα να δίνει σε κάθε γεωγραφική ενότητα σημασία ανάλογη προς την έκτασή της. Γι' αυτό, όταν θέλουμε, εκτός από τα διάφορα επίπεδα του ποσοστού, να τονίσουμε και το απόλυτο μέγεθος που εκφράζει αυτό το ποσοστό (ή το απόλυτο μέγεθος στο οποίο αναφέρεται), χρησιμοποιούμε μιαν άλλη μέθοδο, όπου κάθε γεωγραφική ενότητα παριστάνεται από έναν κύκλο (ή τετράγωνο ή κάποιο άλλο σχήμα) με επιφάνεια ανάλογη προς το συγκεκριμένο απόλυτο μέγεθος. Όμως η ενλόγω μέθοδος παρουσιάζει το βασικό ελάττωμα να διασπά τη γεωγραφική συνέχεια, και γι' αυτό πρέπει να χρησιμοποιείται με πολλή προσοχή και μόνο στις περιπτώσεις που είναι απόλυτα αναγκαίο (βλ., για παράδειγμα, Σχ. 5.8.).

Όταν χρησιμοποιούμε μια κλίμακα τόνων για τη χαρτογράφηση μιας αναλογίας, σε κάθε βαθμίδα της κλίμακας αντιστοιχεί μια τάξη τιμών της αναλογίας. Οι βαθμίδες της κλίμακας δεν πρέπει να είναι περισσότερες από 7 (ή σε εξαιρετικές περιπτώσεις από 9), γιατί αν ο αριθμός των βαθμίδων είναι μεγαλύτερος, το ανθρώπινο μάτι δυσκολεύεται να εντοπίσει τις μεταξύ τους διαφορές. Κατά κανόνα χρησιμοποιούμε κλίμακες με 5 ή 7 βαθμίδες, και σπανιότερα με 4 ή 6.

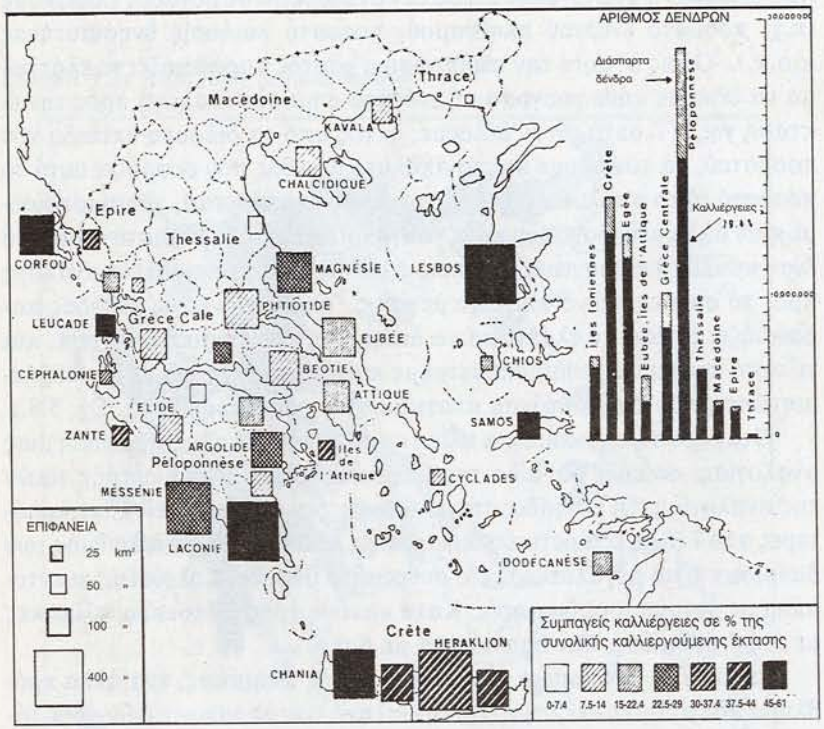
Εκτός από τον αριθμό των βαθμίδων της κλίμακας, ένα άλλο πρόβλημα που αντιμετωπίζουμε είναι η επιλογή των ακραίων τιμών κάθε τάξης. Για το πρόβλημα αυτό δεν υπάρχει ιδανική λύση, αλλά διάφοροι κανόνες λίγο-πολύ αντιφατικοί μεταξύ τους:

1) Τάξεις με ίσο πλάτος: Η μέθοδος αυτή προσφέρεται όταν οι τιμές της αναλογίας είναι περίπου ομαλά κατανεμημένες, και χρησιμοποιείται όταν θέλουμε να τονιστεί η διασπορά του φαινομένου στο χώρο.

2) Ίσος αριθμός γεωγραφικών ενοτήτων σε κάθε τάξη: Η μέθοδος αυτή προσφέρεται όταν θέλουμε να συγκρίνουμε διάφορους χάρτες μεταξύ τους, γιατί μας επιτρέπει να εντοπίσουμε αμέσως την κατάταξη μιας γεωγραφικής ενότητας ως προς διάφορα μεγέθη.

3) Αν η κατανομή των τιμών του μεγέθους που χαρτογραφούμε παρουσιάζει μερικές ομαδοποιήσεις, διαλέγουμε τις ακραίες τιμές κάθε τά-

ΣΧΗΜΑ 5.8. *Συμπαγής καλλιέργεια της ελιάς το 1967 (ανά νομό).*

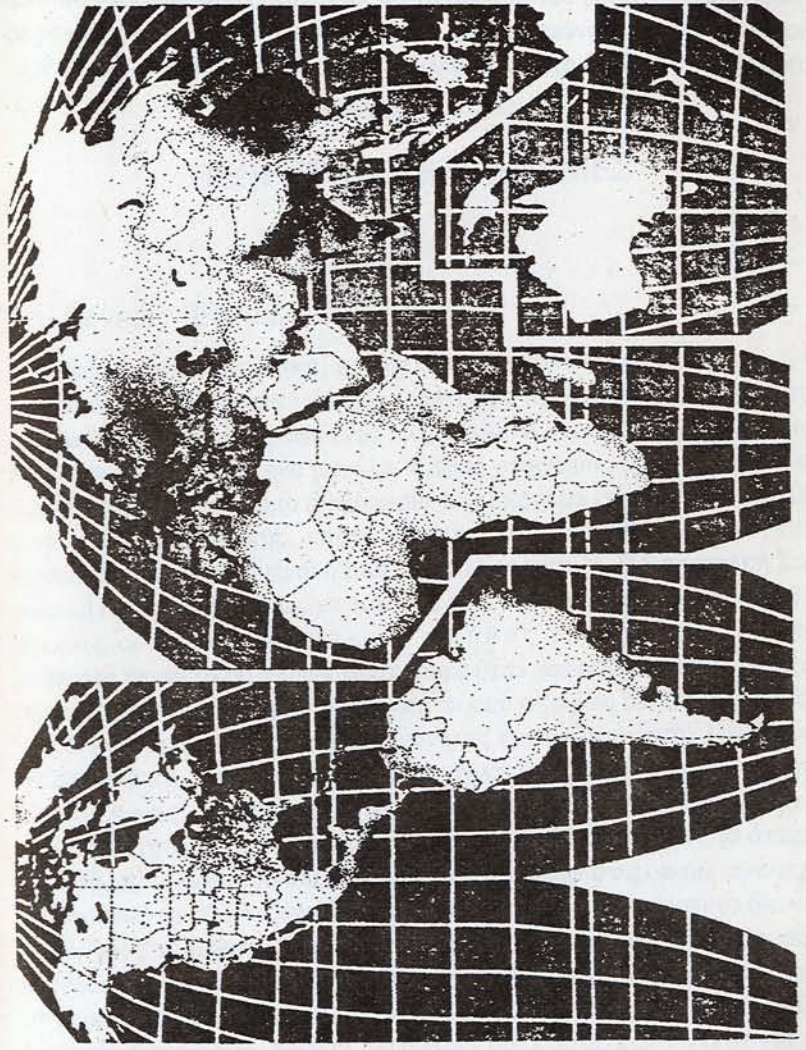


ΠΗΓΗ: KOLODNY, *ό.π.* (τ. 3, Διάγρ. Β.6).

ξης, έτσι ώστε οι γεωγραφικές ενότητες που ανήκουν σε ίδια ομάδα ν' αντιστοιχούν και σε ίδια βαθμίδα της κλίμακας. Η μέθοδος αυτή προσφέρει μια ιδιαίτερα πιστή απεικόνιση του μεγέθους που χαρτογραφείται, αλλά δυσκολεύει τη σύγκριση περισσότερων χαρτών μεταξύ τους.

(γ΄) Τέλος, μια άλλη μέθοδος για την παρουσίαση της γεωγραφικής κατανομής ενός πληθυσμού είναι η χαρτογράφηση με σημεία.

ΣΧΗΜΑ 5.9. *Κατανομή του πληθυσμού της γης (1 σημείο = 100.000 κάτοικοι).*



ΠΗΓΗ: G. CALOT, *Cours de statistique descriptive* (εκδ. Dunod, Παρίσι, 1973, σ. 234).

Στο εσωτερικό κάθε γεωγραφικής ενότητας τοποθετούμε έναν αριθμό σημείων ανάλογο προς τον πληθυσμό της (ή σημεία διαφορετικών διαστάσεων). Το μόνο μειονέκτημα της μεθόδου είναι ότι η κατασκευή ενός τέτοιου χάρτη είναι ιδιαίτερα δύσκολη (βλ., για παράδειγμα, Σχ. 5.9.).

ΚΕΦΑΛΑΙΟ ΕΚΤΟ

ΘΕΩΡΗΤΙΚΕΣ ΚΑΤΑΝΟΜΕΣ

6.1. *Γενικές Έννοιες*

Η Περιγραφική Στατιστική, με την οποία ασχοληθήκαμε ως τώρα, δεν αποτελεί μία μαθηματική θεωρία, αλλά μας προσφέρει τα απαραίτητα εργαλεία για την περιγραφή ενός πληθυσμού ο οποίος, αν και κατά κανόνα αποτελείται από ένα μεγάλο πλήθος στοιχείων, είναι οπωσδήποτε πεπερασμένος. Η μαθηματική θεωρία που αποτελεί τη γενίκευση της Περιγραφικής Στατιστικής, και ταυτόχρονα της προσφέρει την αναγκαία μαθηματική υποδομή, είναι η Επαγωγική Στατιστική. Η Επαγωγική Στατιστική ασχολείται κυρίως με υποθετικούς (θεωρητικούς) πληθυσμούς οι οποίοι έχουν άπειρο πλήθος στοιχείων. Ένα παράδειγμα υποθετικού πληθυσμού έχουμε όταν θεωρούμε ότι ένα πείραμα επαναλαμβάνεται άπειρες φορές «κάτω από ακριβώς όμοιες συνθήκες».

Στην έννοια της σχετικής συχνότητας που ορίσαμε στην Περιγραφική Στατιστική, αντιστοιχεί στην Επαγωγική Στατιστική η έννοια της πιθανότητας. Η πιθανότητα μπορεί να θεωρηθεί ως το όριο της σχετικής συχνότητας όταν το μέγεθος του πληθυσμού τείνει προς το άπειρο, δηλαδή ως το όριο του πηλίκου των περιπτώσεων στις οποίες αναφερόμαστε προς το σύνολο των δυνατών περιπτώσεων, όταν αυτό το δεύτερο τείνει προς το άπειρο. Είναι φανερό ότι οι πιθανότητες —που αφορούν πάντα άπειρους πληθυσμούς— δεν είναι δυνατόν να υπολογιστούν με εμπειρικές μεθόδους αλλά μόνο να εκτιμηθούν· στην πράξη έχουμε μόνο σχετικές συχνότητες, γιατί οι πληθυσμοί που εξετάζουμε είναι πάντα πεπερασμένοι.

Αν υποθέσουμε, π.χ., ότι ρίχνουμε ένα νόμισμα πολλές φορές έτσι ώστε όλες οι ρίψεις να γίνονται «κάτω από ακριβώς όμοιες συνθήκες», θα παρατηρήσουμε πως όσο μεγαλώνει ο αριθμός των ρίψεων τόσο και το πηλίκο του αριθμού των «επιτυχιών» (δηλαδή των περιπτώσεων στις οποίες αναφερόμαστε, π.χ. κορόνα ή γράμματα) πρός τον συνολικό αριθμό των ρίψεων σταθεροποιείται γύρω από έναν αριθμό. Ο αριθμός αυτός αποτελεί μια εκτίμηση του ορίου στο οποίο τείνει η σχετική συχνότητα, δηλαδή αποτελεί μια εκτίμηση της πιθανότητας.

Όπως στη σχετική συχνότητα αντιστοιχεί η πιθανότητα, στις κατανομές συχνοτήτων που μελετάει η Περιγραφική Στατιστική —και που γι' αυτό λέγονται και *εμπειρικές* ή *πειραματικές κατανομές*—, αντιστοιχούν στην Επαγωγική Στατιστική οι θεωρητικές κατανομές. Υπάρχουν πολλές θεωρητικές κατανομές, οι οποίες μας χρησιμεύουν ως (μαθηματικά) πρότυπα για τις πειραματικές κατανομές. Οι θεωρητικές κατανομές χωρίζονται σε δύο κατηγορίες: τις *ασυνεχείς*, που αντιστοιχούν στις ασυνεχείς στατιστικές μεταβλητές, και τις *συνεχείς*, που αντιστοιχούν στις συνεχείς στατιστικές μεταβλητές.

Η αντιστοιχία μεταξύ θεωρητικών και πειραματικών ασυνεχών κατανομών δεν παρουσιάζει ιδιαίτερες δυσκολίες: μία ασυνεχής θεωρητική κατανομή ορίζεται από τις τιμές της (θεωρητικής) ασυνεχούς στατιστικής μεταβλητής και από τις πιθανότητες που αντιστοιχούν σε καθεμιά από αυτές τις τιμές.

Αντίθετα, η αντιστοιχία μεταξύ θεωρητικών και πειραματικών συνεχών κατανομών είναι πιο δύσκολη. Όπως είδαμε στην §3., οι τάξεις μιας πειραματικής συνεχούς στατιστικής μεταβλητής είναι διαστήματα με άκρα τα e_i ($i=0,1, \ldots, \kappa$), και σε κάθε τάξη αντιστοιχεί μια σχετική συχνότητα f_i. Αν διαμερίζουμε αδιάκοπα τα διαστήματα (e_{i-1}, e_i) σε υποδιαστήματα μικρότερου πλάτους, θα καταλήξουμε οπωσδήποτε σε μερικά υποδιαστήματα των οποίων η απόλυτη (επομένως και η σχετική) συχνότητα θα είναι μηδέν. Πραγματικά, τα διαστήματα των οποίων η απόλυτη συχνότητα είναι διάφορη από το μηδέν, δεν μπορεί να είναι περισσότερα από το μέγεθος του πληθυσμού. Αντίθετα, για μια συνεχή θεωρητική κατανομή, η οποία όπως είπαμε αφορά πάντα έναν άπειρο πληθυσμό, υποθέτουμε ότι η διαμέριση μιας τάξης σε υποδιαστήματα, των οποίων

το πλάτος τείνει προς το μηδέν, μπορεί να συνεχίζεται αδιάκοπα χωρίς οι αντίστοιχες σχετικές συχνότητες να μηδενίζονται: η σχετική συχνότητα που αντιστοιχεί σ' ένα οποιοδήποτε διάστημα $(x, x+\Delta x$ —ας τη συμβολίσουμε με $f_{(x,\, x+\Delta x)}$— θα είναι μηδέν μόνον όταν το Δx είναι μηδέν. Όσο το Δx, οσοδήποτε μικρό κι αν είναι, θα παραμένει διάφορο από το μηδέν, η σχετική συχνότητα $f_{(x,\, x+\Delta x)}$, οσοδήποτε μικρή κι αν είναι, θα παραμένει κι αυτή οπωσδήποτε διάφορη από το μηδέν. Έτσι, όπως μια συνεχής πειραματική κατανομή συχνοτήτων παριστάνεται από την πολυγωνική γραμμή συχνότητας και το ιστόγραμμα, μια συνεχής θεωρητική κατανομή συχνοτήτων παριστάνεται από μια συνεχή καμπύλη —που ονομάζεται *καμπύλη συχνότητας*— της οποίας το ύψος σ' ένα σημείο x εκφράζει το όριο, στο σημείο x της σχετικής σχνότητας ανά μονάδα πλάτους, δηλαδή είναι ίσο με $\lim_{\Delta x \to 0} \frac{f_{(x,\, x+\Delta x)}}{\Delta x}$. Η καμπύλη συχνότητας μιας συνεχούς θεωρητικής κατανομής προσδιορίζει πλήρως τη θεωρητική κατανομή, και κατά κανόνα ορίζεται από έναν μαθηματικό τύπο. Με βάση την καμπύλη συχνότητας μπορούμε επίσης να υπολογίσουμε τις διάφορες παραμέτρους κεντρικής τάσης και διασποράς της θεωρητικής κατανομής, όπως π.χ. τον μέσο όρο και την τυπική απόκλιση. Οι έννοιες αυτές, για μια συνεχή θεωρητική κατανομή, έχουν σημασία αντίστοιχη μ' εκείνην που έχουν και για μια πειραματική συνεχή κατανομή.

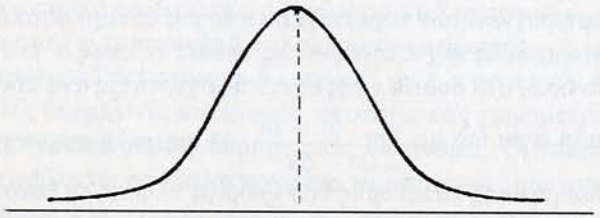
Η μελέτη των θεωρητικών κατανομών αποτελεί, όπως είπαμε, θέμα της Επαγωγικής Στατιστικής, και απαιτεί μια σειρά από μαθηματικές γνώσεις. Γι' αυτό εδώ θα περιοριστούμε μόνο σε μια πρώτη, πολύ γενική, παρουσίαση μιας μόνο συνεχούς θεωρητικής κατανομής, της κανονικής κατανομής, η οποία, εκτός από το ιδιαίτερο θεωρητικό ενδιαφέρον που παρουσιάζει, χρησιμοποιείται πολύ συχνά στις πρακτικές εφαρμογές, γιατί αρκετές πειραματικές κατανομές προσεγγίζονται ικανοποιητικά από την κανονική κατανομή.

6.2. *Η Κανονική Κατανομή*

6.2.1. *Ορισμοί και Ιδιότητες*

Κανονική κατανομή ονομάζεται μια οικογένεια θεωρητικών κατανομών των οποίων η καμπύλη συχνότητας είναι μια ειδική συμμετρική καμπύλη —που λέγεται *κωδωνοειδής*—, κι έχει την εξής γενική μορφή:

ΣΧΗΜΑ 6.1. *Γενική μορφή της καμπύλης συχνότητας της κανονικής κατανομής.*



Η καμπύλη του Σχήματος 6.1. ορίζεται μαθηματικά από τη συνάρτηση:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2-\mu}{2\sigma^2}},$$

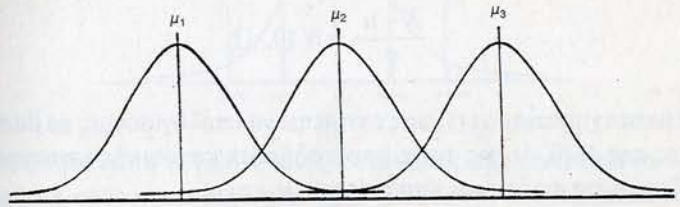
όπου μ είναι ο μέσος όρος και σ η τυπική απόκλιση της κανονικής κατανομής. Υπάρχουν λοιπόν άπειρες κανονικές κατανομές —ανάλογα με τις τιμές που παίρνουν οι παράμετροι μ και σ— και, αντίστροφα, μια κανονική κατανομή προσδιορίζεται εντελώς από τη στιγμή που γνωρίζουμε τις τιμές που έχουν οι παράμετροι μ και σ.

Οι πιο βασικές ιδιότητες της κανονικής κατανομής είναι οι εξής:

- Η καμπύλη συχνότητας της κανονικής κατανομής είναι συμμετρική ως προς την κατακόρυφο που τέμνει τον άξονα των x στο σημείο μ (μέσος όρος της κατανομής).
- Η προηγούμενη ιδιότητα έχει ως συνέπεια ότι τα τρία χαρακτηριστικά κεντρικής τάσης (μέσος όρος, διάμεσος, τύπος) της κανονικής κατανομής ταυτίζονται.

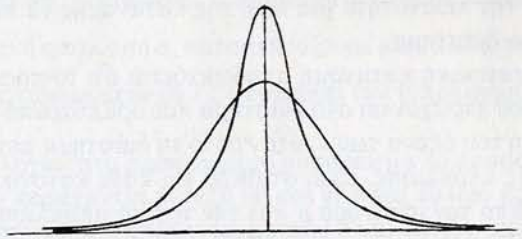
- Οι καμπύλες συχνότητας των κανονικών κατανομών που έχουν την ίδια τυπική απόκλιση ανάγονται η μια στην άλλη με παράλληλη μετατόπιση (βλ. Σχ. 6.2.).

ΣΧΗΜΑ 6.2. *Σύγκριση κανονικών κατανομών με ίδια τυπική απόκλιση και διαφορετικούς μέσους όρους.*



- Ανάλογα με την τιμή της τυπικής απόκλισης η καμπύλη συχνότητας είναι περισσότερο ή λιγότερο διασπαρμένη (βλ. Σχ. 6.3.).

ΣΧΗΜΑ 6.3. *Σύγκριση δύο κανονικών κατανομών με ίδιο μέσο όρο και διαφορετική τυπική απόκλιση.*



Μια ειδική μορφή της κανονικής κατανομής, την οποία χρησιμοποιούμε πολύ συχνά και οι χαρακτηριστικότερες τιμές της οποίας βρίσκονται σε πίνακες (βλ. Παράρτημα), είναι η *τυποποιημένη κανονική κατανομή*. Ο μέσος όρος μ (επομένως και η διάμεσος και ο τύπος) της τυποποιημένης κανονικής κατανομής είναι ίσος με 0, και η τυπική απόκλιση σ είναι ίση με 1. Την τυποποιημένη κανονική κατανομή τη συμβο-

λίζουμε με $N(0, 1)$, ενώ για μια κανονική κατανομή με μέσο όρο μ και τυπική απόκλιση σ χρησιμοποιούμε το συμβολισμό $N(\mu, \sigma^2)$. Η σημασία της $N(0, 1)$ προέρχεται από το γεγονός ότι για οποιαδήποτε κανονική κατανομή X, με μέσο όρο μ και τυπική απόκλιση σ, η κατανομή $\frac{X-\mu}{\sigma}$ είναι η τυποποιημένη κανονική κατανομή, δηλαδή:

$$\frac{X-\mu}{\sigma} = N(0, 1) \qquad (1)$$

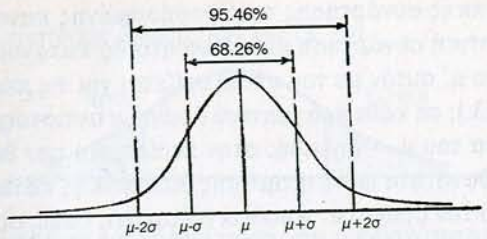
Ο μετασχηματισμός (1) μας επιτρέπει να υπολογίσουμε, με βάση τους πίνακες της $N(0, 1)$, τις τιμές οποιασδήποτε κανονικής κατανομής αν γνωρίζουμε τις τιμές των παραμέτρων μ και σ.

6.2.2. *Αθροιστική Συνάρτηση και Πίνακες*

Είναι πολύ συχνά χρήσιμο (βλ. π.χ. §7., σχετικά με τον έλεγχο υποθέσεων) να μπορούμε να προσδιορίσουμε το ποσοστό των περιπτώσεων που περιέχονται σ' ένα δεδομένο διάστημα τιμών μιας θεωρητικής κατανομής, δηλαδή την πιθανότητα μια τιμή της κατανομής να περιέχεται στο συγκεκριμένο διάστημα.

Για την κανονική κατανομή αποδεικνύεται ότι το ποσοστό των περιπτώσεων που περιέχονται στο διάστημα που ορίζεται από τον μέσο όρο μ και μια τιμή του άξονα των x, όταν αυτό το διάστημα μετριέται σε μονάδες τυπικής απόκλισης, είναι σταθερό για κάθε κανονική κατανομή, ανεξάρτητα από τον μέσο όρο μ και την τυπική απόκλιση σ αυτής της κατανομής. Έτσι, π.χ., για οποιαδήποτε κανονική κατανομή στο διάστημα $(\mu, \mu+\sigma)$ περιέρχονται τα 34,13% των περιπτώσεων, και στο διάστημα $(\mu-\sigma, \mu+\sigma)$ ο διπλάσιος αριθμός περιπτώσεων, δηλαδή τα 68,26% (βλ. Σχ. 6.4.).

ΣΧΗΜΑ 6.4.



Η ιδιότητα αυτή της κανονικής κατανομής μάς επιτρέπει να υπολογίσουμε, για κάθε κανονική κατανομή, τι ποσοστό των περιπτώσεων περιέχεται σ' ένα οποιοδήποτε δεδομένο διάστημα. Ας υποθέσουμε π.χ. ότι έχουμε μια κανονική κατανομή με μέσο όρο 50 και τυπική απόκλιση 10. Για να βρούμε τι ποσοστό των περιπτώσεων περιέχεται, π.χ., μεταξύ του 50 και του 65, θα πρέπει να καθορίσουμε σε πόσες μονάδες τυπικής απόκλισης αντιστοιχεί η διαφορά των δύο αυτών τιμών, δηλαδή θα πρέπει να διαιρέσουμε τη διαφορά με την τυπική απόκλιση $\left(\frac{65-50}{10} = 1{,}5\right)$. Αυτό όμως σημαίνει ότι, ξεκινώντας από μια κανονική κατανομή X με μέσο όρο μ και τυπική απόκλιση σ, κατασκευάζουμε μιαν άλλη, τη $Z = \frac{X-\mu}{\sigma}$, η οποία, όπως είδαμε στην §6.1., ακολουθεί την τυποποιημένη κανονική κατανομή.

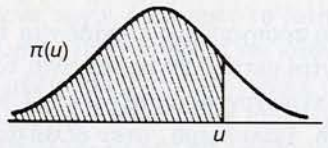
Μ' άλλα λόγια, στο προηγούμενο παράδειγμα το ποσοστό των περιπτώσεων που περιέχονται μεταξύ 50 και 65 είναι το ίδιο με το ποσοστό των περιπτώσεων που περιέχονται μεταξύ 0 και 1,5 για την τυποποιημένη κανονική κατανομή. Γενικότερα, όταν θέλουμε να προσδιορίσουμε την πιθανότητα ώστε οι τιμές μιας κανονικής κατανομής να είναι μικρότερες από έναν συγκεκριμένο αριθμό ή να περιέχονται σ' ένα συγκεκριμένο διάστημα, αρκεί να προσδιορίσουμε την πιθανότητα ώστε οι τιμές της τυποποιημένης κανονικής κατανομής να είναι μικρότερες από έναν (κατάλληλα προσδιορισμένο) αριθμό ή να περιέχονται σ' ένα (κατάλληλα προσδιορισμένο) διάστημα.

Η πιθανότητα ώστε η τιμή της τυποποιημένης κανονικής κατανομής να είναι μικρότερη από έναν αριθμό u είναι ίση με την τιμή, στο σημείο u, της αθροιστικής συνάρτησης της τυποποιημένης κανονικής κατανομής. Η αθροιστική συνάρτηση μιας θεωρητικής κατανομής ορίζεται με τρόπο ανάλογο μ' αυτόν με τον οποίο ορίζεται για τις πειραματικές κατανομές (βλ. §3.): σε κάθε πραγματικό αριθμό u αντιστοιχεί η ολική σχετική συχνότητα του u — δηλαδή, στην περίπτωση των θεωρητικών κατανομών, η πιθανότητα ώστε η τιμή της θεωρητικής κατανομής να είναι μικρότερη από τον αριθμό u. Αποδεικνύεται ότι, όπως συμβαίνει και με τις συνεχείς πειραματικές κατανομές (βλ. §3.2.), η τιμή της αθροιστικής συνάρτησης στο σημείο u είναι ίση με το εμβαδόν που περικλείεται από την αντίστοιχη καμπύλη συχνότητας και τον άξονα των x από το $-\infty$ ως το u (βλ. Σχ. 6.5.).

Η αθροιστική συνάρτηση της τυποποιημένης κανονικής κατανομής συμβολίζεται με $\Pi(u)$, και οι χαρακτηριστικότερες τιμές της για $u \geqslant 0$ δίνονται από τον πίνακα που βρίσκεται στο τέλος του βιβλίου.

Σύμφωνα με όσα είπαμε παραπάνω, η πιθανότητα ώστε η τιμή της τυποποιημένης κανονικής κατανομής να είναι μικρότερη από έναν αριθμό μ (δηλαδή η τιμή της αθροιστικής της συνάρτησης στο σημείο u) είναι ίση με το εμβαδόν που περικλείεται από την αντίστοιχη κωδωνοειδή καμπύλη και τον άξονα των x από το $-\infty$ ως το u (βλ. Σχ. 6.5.).

ΣΧΗΜΑ 6.5. *Αθροιστική συνάρτηση της τυποποιημένης κανονικής κατανομής.*

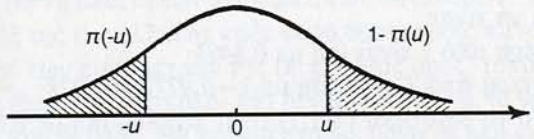


Αντίστοιχα, η πιθανότητα ώστε η τιμή της τυποποιημένης κανονικής κατανομής να είναι μεγαλύτερη από u, είναι ίση με $1-\Pi(u)$ (βλ. και Σχ. 6.6.).

Όταν το u είναι αρνητικό, για να προσδιορίσουμε την τιμή της αθροιστικής συνάρτησης $\Pi(u)$ (στον πίνακα υπάρχουν μόνο οι τιμές της για $u \geqslant 0$), επειδή η καμπύλη συχνότητας είναι συμμετρική ως προς την κα-

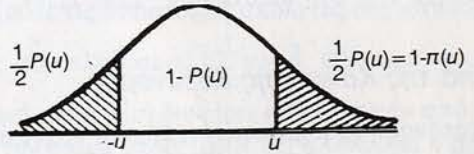
τακόρυφο που περνάει από το 0, εφαρμόζουμε τον τύπο $\Pi(-u) = 1 - \Pi(u)$, που μια γεωμετρική του εξήγηση δίνει το Σχήμα 6.5.

ΣΧΗΜΑ 6.6. *Αντιστοιχία μεταξύ $\Pi(u)$ και $\Pi(-u)$.*



Από τα παραπάνω προκύπτει επίσης ότι η πιθανότητα ώστε η τιμή της τυποποιημένης κανονικής κατανομής να βρίσκεται έξω από ένα διάστημα $(-u, u)$ είναι ίση με $P(u) = 2[1 - \Pi(u)]$, όπως φαίνεται και στο Σχήμα 6.7.

ΣΧΗΜΑ 6.7. *Γεωμετρική σημασία της συνάρτησης $P(u)$.*



Αντίστοιχα, η πιθανότητα ώστε η τιμή της τυποποιημένης κανονικής κατανομής να βρίσκεται μέσα στο διάστημα $(-u, u)$, είναι ίση με $1 - P(u)$ (βλ. και Σχ. 6.7.).

Για να μπορούμε να προσδιορίζουμε με μεγαλύτερη ευκολία την πιθανότητα ώστε η τιμή της τυποποιημένης κανονικής κατανομής να βρίσκεται μέσα ή έξω από ένα διάστημα $(-u, u)$, στο τέλος του βιβλίου υπάρχει ένας πίνακας με τις χαρακτηριστικότερες τιμές της συνάρτησης $P(u)$.

Τόσο ο πίνακας με τις τιμές της συνάρτησης $\Pi(u)$ όσο και ο πίνακας με τις τιμές της συνάρτησης $P(u)$, μπορούν να χρησιμοποιηθούν κατά δύο τρόπους:

(α΄) Όταν θέλουμε να προσδιορίσουμε την πιθανότητα ώστε η τιμή της τυποποιημένης κανονικής κατανομής να βρίσκεται έξω (ή μέσα) από ένα δεδομένο διάστημα $(-u, u)$ ή να είναι μικρότερη (ή μεγαλύτερη) από έναν δεδομένο αριθμό u.

(β΄) Όταν θέλουμε να προσδιορίσουμε, με δεδομένη πιθανότητα, το διάστημα μέσα στο οποίο θα βρίσκεται η τιμή της τυποποιημένης κανονικής κατανομής.

Έτσι, π.χ., η πιθανότητα ώστε η τιμή της τυποποιημένης κανονικής κατανομής να είναι:

— μικρότερη από 1 είναι ίση με 0,8413·
— μεγαλύτερη από 2 είναι ίση με $1 - 0{,}9772 = 0{,}0228$·
— έξω από το διάστημα $(-1{,}1)$ είναι περίπου ίση με 0,3174·
— μέσα στο διάστημα $(-2{,}2)$ είναι περίπου ίση με 0,9546.

Αντίστροφα, οι τιμές της τυποποιημένης κανονικής κατανομής βρίσκονται:

— με πιθανότητα 0,90 στο διάστημα $(-1{,}64,\ 1{,}64)$·
— με πιθανότητα 0,95 στο διάστημα $(-1{,}96,\ 1{,}96)$·
— με πιθανότητα 0,99 στο διάστημα $(-2{,}58,\ 2{,}58)$.

6.2.3. *Σημασία της Κανονικής Κατανομής*

Η κανονική κατανομή χρησιμοποιείται αρκετά συχνά ως μαθηματικό υπόδειγμα για την περιγραφή διαφόρων φαινομένων (π.χ. κατανομή των ατόμων ως προς το ύψος κ.τ.λ.). Μία από τις σημαντικότερες όμως εφαρμογές της είναι η χρησιμοποίησή της για τον προσδιορισμό των «τυχαίων λαθών». Πράγματι, θεωρούμε ότι τα «τυχαία λάθη» (ή οι «τυχαίες αποκλίσεις») κατανέμονται γύρω από την «πραγματική τιμή» σύμφωνα με την κανονική κατανομή. Επομένως, η χρήση της κανονικής κατανομής μάς επιτρέπει να συμπεράνουμε αν τα λάθη που εξετάζουμε οφείλονται πράγματι στην τύχη ή σε κάποια άλλη αιτία.

Οι εφαρμογές της κανονικής κατανομής που αναφέραμε πιο πάνω βασίζονται στην εξής πολύ σημαντική ιδιότητα της κανονικής κατανομής, που αποδεικνύεται στη θεωρία πιθανοτήτων (κεντρικό οριακό θεώρημα): αν μία μεταβλητή X αντιπροσωπεύει το αποτέλεσμα ενός πολύ μεγάλου αριθμού από ανεξάρτητες αιτίες με προστιθέμενες επιπτώσεις, κι αν η επίπτωση κάθε αιτίας ξεχωριστά είναι αμελητέα σε σχέση με το σύνολο των επιπτώσεων, τότε η μεταβλητή X ακολουθεί την κανονική κατανομή.

6.3. *Απόσταση μεταξύ Θεωρητικής και Πειραματικής Κατανομής*

Οι θεωρητικές κατανομές χρησιμοποιούνται, όπως είπαμε, ως μαθηματικά πρότυπα για τη μελέτη των πειραματικών κατανομών. Το πρόβλημα της επιλογής της κατάλληλης κάθε φορά θεωρητικής κατανομής και του καθορισμού των παραμέτρων της (π.χ. μέσος όρος, τυπική απόκλιση κλπ.) είναι ιδιαίτερα δύσκολο, και δεν μπορούμε να το θίξουμε στα πλαίσια αυτού του βιβλίου. Θα περιοριστούμε μόνο σε μερικές πολύ γενικές παρατηρήσεις σχετικά με τους όρους κάτω από τους οποίους μπορούμε να θεωρήσουμε ότι μια θεωρητική κατανομή αποτελεί ικανοποιητική προσέγγιση μιας πειραματικής κατανομής.

Ας υποθέσουμε ότι έχουμε μια στατιστική μεταβλητή Π που παρουσιάζει K τάξεις, τις οποίες συμβολίζουμε με $X_1, \ldots, X_i, \ldots, X_\kappa$. Με $\nu_1, \ldots, \nu_i, \ldots, \nu_\kappa$ συμβολίζουμε τις αντίστοιχες απόλυτες συχνότητες, και με ν το συνολικό μέγεθος του πληθυσμού $\left(\sum_{i=1}^{\kappa} \nu = \nu\right)$. Θέλουμε να ελέγξουμε αν η πειραματική αυτή κατανομή διαφέρει «σημαντικά» από μια θεωρητική κατανομή Θ, την οποία υποθέτουμε ότι ακολουθεί η στατιστική μεταβλητή Π.

Συμβολίζουμε με $\rho_1, \ldots, \rho_i, \ldots, \rho_\kappa$ τις πιθανότητες που αντιστοιχούν, σύμφωνα με τη θεωρητική κατανομή, στις K τάξεις $X_1, \ldots, X_i, \ldots, X_\kappa$ της στατιστικής μεταβλητής. Επειδή $\sum_{i=1}^{\kappa} \rho_i = 1$, θα έχουμε $\sum_{i=1}^{\kappa} \nu\rho_i = \nu$. Οι αριθμοί $\nu\rho_1, \ldots, \nu\rho_i, \ldots, \nu\rho_\kappa$ ονομάζονται *θεωρητικές απόλυτες συχνότητες*.

Τάξεις	*Πειραματικές απόλυτες συχνότητες*	*Θεωρητικές απόλυτες συχνότητες*
X_1	ν_1	$\nu\rho_1$
⋮	⋮	⋮
X_i	ν_i	$\nu\rho_i$
⋮	⋮	⋮
X_κ	ν_κ	$\nu\rho_\kappa$

Ως απόσταση[1] μεταξύ της πειραματικής κατανομής Π και της θεωρητικής κατανομής Θ ορίζουμε το μέγεθος:

$$\Delta = \sum_{i=1}^{\kappa} \frac{(\nu_i - \nu\rho_i)^2}{\nu\rho_i},$$

ή συμβολικά:

$$\Delta = \sum_{i=1}^{\kappa} \frac{(\Pi_i - \Theta_i)^2}{\Theta_i}.$$

Αποδεικνύεται ότι, οποιαδήποτε κι αν ειναι η θεωρητική κατανομή Θ, αν είναι αληθινή η υπόθεση ότι η στατιστική μεταβλητή Π ακολουθεί τη θεωρητική κατανομή Θ, τότε η απόσταση που ορίσαμε πιο πάνω ακολουθεί μια γνωστή θεωρητική κατανομή που ονομάζεται *κατανομή του* X^2 *με* μ *βαθμούς ελευθερίας*.[2]

Ο αριθμός μ των βαθμών ελευθερίας είναι μια παράμετρος που εμπειρικά εκφράζει τον αριθμό των τάξεων της θεωρητικής κατανομής για τις οποίες θα μπορούσαμε να ορίσουμε αυθαίρετα τις απόλυτες συχνότητες. Αν η στατιστική μεταβλητή παρουσιάζει K τάξεις, και για να προσδιορίσουμε τη θεωρητική κατανομή Θ εκτιμήσαμε, με βάση τα δεδομένα των παρατηρήσεων, ρ παραμέτρους, ο αριθμός των βαθμών ελευθερίας είναι ίσος με $\mu = K - \rho - 1$.

Η κατανομή του x^2 διαφέρει ανάλογα με τον αριθμό των βαθμών ελευθερίας. Ο πίνακας που υπάρχει στο τέλος του βιβλίου μάς δίνει για κάθε βαθμό ελευθερίας ($\mu < 30$) τις πιο χαρακτηριστικές τιμές της κατανομής του x^2.

Για να ελέγξουμε λοιπόν αν η πειραματική κατανομή Π διαφέρει «σημαντικά» από τη θεωρητική κατανομή Θ, υπολογίζουμε την απόσταση Δ και βρίσκουμε από τους πίνακες ποια είναι η πιθανότητα η κατανομή

1. Στην πραγματικότητα, το μέγεθος Δ δεν ανταποκρίνεται στον μαθηματικό ορισμό της απόστασης, γιατί $\Delta(\Pi_i, \Theta_i) \neq \Delta(\Theta_i, \Pi_i)$. Είναι όμως χρήσιμο να ονομάσουμε το Δ *απόσταση*, γιατί αυτό ανταποκρίνεται στη σημασία με την οποία χρησιμοποιείται.

2. Αυτό σημαίνει ότι αν υλοποιήσουμε άπειρες φορές τη θεωρητική κατανομή Θ, οι διάφορες τιμές που παίρνει η απόσταση ανάμεσα στη Θ και τις πειραματικές υλοποιήσεις της ακολουθούν την κατανομή του x^2.

x^2, για τον αριθμό των βαθμών ελευθερίας που μας ενδιαφέρει, να ξεπερνάει την απόσταση Δ που υπολογίσαμε. Αν η πιθανότητα αυτή είναι μικρή, δηλαδή μικρότερη από ένα όριο που έχουμε ορίσει εκ των προτέρων (π.χ. 0.05 ή 0.01), θεωρούμε ότι η πειραματική κατανομή Π διαφέρει σημαντικά από τη θεωρητική κατανομή Θ. Αν, αντίθετα, η πιθανότητα είναι μεγαλύτερη απ' αυτό το όριο, θεωρούμε ότι οι διαφορές μεταξύ πειραματικής και θεωρητικής κατανομής μπορεί να οφείλονται στην τύχη, κι επομένως η υπόθεση ότι οι παρατηρήσεις ακολουθούν τη θεωρητική κατανομή Θ είναι α π ο δ ε κ τ ή.

Παράδειγμα: Ας υποθέσουμε π.χ. ότι σε μια χώρα, στη διάρκεια ενός χρόνου, γεννήθηκαν 100.000 παιδιά, από τα οποία τα 50.500 ήταν κορίτσια και τα 49.500 αγόρια. Είναι λογικό να υποθέσουμε ότι οι γεννήσεις αυτές ακολουθούν μια θεωρητική κατανομή σύμφωνα με την οποία η πιθανότητα να γεννηθεί ένα κορίτσι και η πιθανότητα να γεννηθεί ένα αγόρι είναι ίση με 0.5;

Αν υπολογίσουμε την απόσταση Δ, έχουμε:

$$\Delta = \frac{(50.000 - 49.500)^2}{50.000} + \frac{(50.000 - 50.500)^2}{50.000} =$$

$$= \frac{(500)^2 + (500)^2}{50.000} = \frac{500.000}{50.000} = 10.$$

Αν η υπόθεσή μας είναι αληθινή, πρέπει η απόσταση Δ ν' ακολουθεί την κατανομή του x^2 με 1 βαθμό ελευθερίας ($\kappa = 2$, $\rho = 0$). Από τον πίνακα των τιμών του x^2 όμως βρίσκουμε ότι για 1 βαθμό ελευθερίας η πιθανότητα ώστε η τιμή του x^2 να ξεπερνάει το 6,635 είναι 0.01. Είμαστε επομένως υποχρεωμένοι να θεωρήσουμε ότι η πειραματική κατανομή των γεννήσεων που παρατηρήθηκαν διαφέρει σημαντικά από τη θεωρητική κατανομή που υποθέσαμε.

Η χρησιμοποίηση του x^2, για να ελέγξουμε την υπόθεση ότι η πειραματική κατανομμή Π ακολουθεί τη θεωρητική κατανομή Θ, αποτελεί ένα παράδειγμα αυτού που ονομάζουμε *στατιστικό έλεγχο υποθέσεων*. Με τον έλεγχο των υποθέσεων θ' ασχοληθούμε αναλυτικότερα στην επόμενη παράγραφο. Πρέπει μόνο εδώ να σημειώσουμε ότι για ένα συγκε-

κριμένο σύνολο παρατηρήσεων είναι πιθανόν να μπορούν να θεωρηθούν ως αποδεκτά μαθηματικά υποδείγματα περισσότερες από μία θεωρητικές κατανομές. Μια υπόθεση αποδεκτή δεν είναι μια υπόθεση αναγκαστικά αληθινή, όπως θα δούμε στο επόμενο Κεφάλαιο. Αντίθετα, μια υπόθεση που κρίνεται μη αποδεκτή είναι κατά πάσα πιθανότητα μια λανθασμένη υπόθεση. Μ' αυτή την έννοια, ο έλεγχος μιας υπόθεσης είναι θετικός (δηλαδή προσφέρει συγκεκριμένη πληροφορία) στην περίπτωση που είναι αρνητικός (δηλαδή όταν οδηγεί στην απόρριψη μιας υπόθεσης).

ΚΕΦΑΛΑΙΟ ΕΒΔΟΜΟ

ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ

7.1. *Γενικές Έννοιες*

Σε πάρα πολλές περιπτώσεις (οικονομικές έρευνες, κοινωνιολογικές έρευνες κ.τ.λ.) επιθυμούμε να διατυπώσουμε συμπεράσματα σχετικά με το σύνολο ενός πληθυσμού, ενώ τα δεδομένα που διαθέτουμε προέρχονται από ένα δείγμα[1] μόνο του πληθυσμού αυτού. Ή, αντίστροφα, γνωρίζοντας μερικά στοιχεία για το σύνολο του πληθυσμού, θέλουμε να ελέγξουμε την αντιπροσωπευτικότητα του δείγματος που έχουμε επιλέξει. Και στις δύο περιπτώσεις ακολουθούμε την ίδια διαδικασία: διατυπώνουμε καταρχήν μια υπόθεση, και στη συνέχεια επιδιώκουμε ν' αποφανθούμε (δηλαδή ν' αποφασίσουμε) αν πρέπει να τη δεχτούμε ή να την απορρίψουμε. Η διαδικασία αυτή λέγεται *έλεγχος* (ή τεστ) *υπόθεσης*. Η θεωρία των τεστ αποτελεί ένα πολύ βασικό τμήμα της Επαγωγικής Στατιστικής, η οποία ονομάζεται εξάλλου Επαγωγική γιατί ακριβώς ασχολείται με τα θεωρητικά εκείνα εργαλεία που μας επιτρέπουν το πέρασμα από το μερικό στο γενικό.

Ένα παράδειγμα ελέγχου υπόθεσης είδαμε στην §6.3. όταν εξετάσαμε κάτω από ποιες προϋποθέσεις μια θεωρητική κατανομή μπορεί να θεωρηθεί ως αποδεκτό πρότυπο μιας πειραματικής κατανομής.

1. Για τις έννοιες *δείγμα*, *δειγματοληπτική κατανομή* κ.τ.λ. που χρησιμοποιούνται σ' αυτή την παράγραφο, βλ. το Κεφάλαιο περί δειγματοληψίας.

Υπάρχουν πολλές μέθοδοι για τον έλεγχο των υποθέσεων, όλες όμως βασίζονται στην παρακάτω γενική διαδικασία:

Πρώτον, διατυπώνουμε την υπόθεση που θέλουμε να ελέγξουμε και την οποία συμβολίζουμε με H_0. Ταυτόχρονα διαλέγουμε και τον τρόπο με τον οποίο θα ελέγξουμε αυτή την υπόθεση, δηλαδή την παράμετρο του δείγματος (π.χ. μέσο όρο, τυπική απόκλιση κ.τ.λ.) που θα υπολογίσουμε και τη θεωρητική κατανομή που ακολουθεί η δειγματοληπτική κατανομή αυτής της παραμέτρου.

Δεύτερον, καθορίζουμε το επίπεδο σημαντικότητας, δηλαδή καθορίζουμε ποια πιθανότητα θεωρούμε ως τη μικρότερη για την οποία δεχόμαστε ότι οι διαφορές μεταξύ δείγματος και πληθυσμού (διαφορές δηλαδή των αντίστοιχων παραμέτρων που ελέγχουμε) οφείλονται στη «φυσική διακύμανση των τυχαίων δειγμάτων», δηλαδή στην ίδια τη δειγματοληψία. Στις κοινωνικές επιστήμες, ως επίπεδο σημαντικότητας επιλέγουμε κατά κανόνα την πιθανότητα 0.05 ή 0.01.

Τρίτον, υπολογίζουμε το στατιστικό τεστ, δηλαδή υπολογίζουμε, από τα δεδομένα του δείγματος, μία ποσότητα η οποία κατανέμεται σύμφωνα με μια γνωστή θεωρητική κατανομή. Ύστερα εκτιμούμε την πιθανότητα να είναι πραγματικά η τιμή που βρήκαμε μια τιμή αυτής της θεωρητικής κατανομής. Φυσικά υπάρχουν πολλές ποσότητες που μπορούν να υπολογιστούν από τα δεδομένα ενός δείγματος, αλλά μόνον λίγες απ' αυτές ακολουθούν γνωστές δειγματοληπτικές κατανομές ώστε να μπορούν να χρησιμοποιηθούν για τον έλεγχο υποθέσεων.

Τέταρτον, έχοντας εκτιμήσει την πιθανότητα τα δεδομένα του δείγματος να προέρχονται από έναν πληθυσμό που επαληθεύει την υπόθεση H_0, παίρνουμε τη στατιστική απόφαση. Αν η πιθανότητα αυτή είναι μικρότερη από το επίπεδο σημαντικότητας που έχουμε επιλέξει, απορρίπτουμε την υπόθεση H_0, γιατί θεωρούμε ότι υπάρχει σημαντική διαφορά ανάμεσα στα δεδομένα του δείγματος και στα χαρακτηριστικά που θα όφειλε να είχε ο πληθυσμός αν επαλήθευε την υπόθεση H_0. Στατιστικά σημαντική διαφορά σημαίνει ότι η πιθανότητα να παρουσιαστεί μια τέτοια διαφορά είναι μικότερη από το επίπεδο σημαντικότητας που έχουμε επιλέξει. Αν, αντίθετα, η πιθανότητα που υπολογίσαμε είναι μεγαλύτερη από το επίπεδο σημαντικότητας που έχουμε ορίσει, θεωρούμε ότι

η υπόθεση H_0 είναι αποδεκτή, χωρίς φυσικά αυτό να σημαίνει ότι είναι και αναγκαστικά αληθινή.

Είναι φανερό ότι στη διαδικασία με την οποία παίρνουμε μια στατιστική απόφαση μπορεί να οδηγηθούμε στην απόρριψη της αρχικής υπόθεσης ενώ αυτή είναι αληθινή. Ένα σφάλμα τέτοιου είδους λέγεται *σφάλμα τύπου Ι*. Αντίστροφα, είναι επίσης δυνατό να θεωρήσουμε αποδεκτή μια υπόθεση η οποία στην πραγματικότητα είναι λανθασμένη. Το σφάλμα αυτού του είδους λέγεται *σφάλμα τύπου ΙΙ*.

Η θεωρία των τεστ είναι εξαιρετικά χρήσιμη για την κοινωνιολογική έρευνα, αλλά ξεφεύγει από τα πλαίσια αυτού του βιβλίου. Θα περιοριστούμε λοιπόν σ' ένα μόνο παράδειγμα, με το οποίο ελπίζουμε να γίνει πιο κατανοητή η διαδικασία με την οποία παίρνεται μια στατιστική απόφαση.

7.2. *Παράδειγμα*

Ας υποθέσουμε ότι ένας ερευνητής πρόκειται να επεξεργαστεί τ' αποτελέσματα μιας δειγματοληπτικής έρευνας που έγινε σχετικά με το εισόδημα των οικογενειών μιας ορισμένης κοινότητας. Ξέρει ότι ο μέσος όρος του οικογενειακού εισοδήματος στη συγκεκριμένη κοινότητα είναι 150.000 δρχ. κι ότι η τυπική απόκλιση είναι 50.000 δρχ. Ο μέσος όρος του οικογενειακού εισοδήματος για τις οικογένειες του δείγματος είναι 165.000 δρχ., και το μέγεθος του δείγματος, δηλαδή ο αριθμός των οικογενειών που ρωτήθηκαν, είναι 100. Είναι μήπως λογικό να θεωρήσει ο ερευνητής ότι η δειγματοληψία έγινε από μη έμπειρους συνεντευκτές, με αποτέλεσμα να υπεραντιπροσωπεύονται οι οικογένειες μεσαίου και υψηλού εισοδήματος; Για να ελέγξει την ορθότητα της δειγματοληψίας, δηλαδή την αντιπροσωπευτικότητα του δείγματος, πρέπει να ακολουθήσει την εξής διαδικασία:

1) *Καθορισμός υποθέσεων*. Ως αρχική υπόθεση H_0 (την οποία ακριβώς θα θέλαμε ν' απορρίψουμε) θεωρούμε την αντιπροσωπευτικότητα του δείγματος, δηλαδή ότι όλες οι οικογένειες της συγκεκριμένης κοινότητας είχαν την ίδια πιθανότητα να επιλεγούν στο δείγμα. Αν η υπόθεση αυτή αληθεύει, μπορούμε να θεωρήσουμε ότι η δειγματοληπτική κατα-

νομή των μέσων όρων των τυχαίων δειγμάτων μεγέθους N είναι μια κανονική κατανομή με μέσο όρο τον μέσο όρο του δείγματος μ και τυπική απόκλιση $\sigma_{\bar{x}}$ το $\frac{1}{\sqrt{N}}$ της τυπικής απόκλισης σ του πληθυσμού (δηλαδή είναι $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$).

Για μέγεθος δείγματος $N > 30$ μπορούμε να θεωρήσουμε την κανονική κατανομή ως το θεωρητικό πρότυπο της δειγματοληπτικής κατανομής των μέσων όρων των τυχαίων δειγμάτων μεγέθους N. Αντίθετα, για $N \leqslant 30$ δεν χρησιμοποιούμε την κανονική κατανομή ως θεωρητικό πρότυπο παρά μόνο αν γνωρίζουμε εκ των προτέρων ότι η προσέγγιση είναι ικανοποιητική. Κατά κανόνα, όταν έχουμε μικρά δείγματα ($N \leqslant 30$) χρησιμοποιούμε άλλες θεωρητικές κατανομές, όπως π.χ. την κατανομή t του Student.

2) *Επιλογή του επιπέδου σημαντικότητας.* Ας υποθέσουμε ότι στη συγκεκριμένη περίπτωση μας αρκεί ένα επίπεδο σημαντικότητας 0.05.

Η επιλογή του επιπέδου σημαντικότητας γίνεται με βάση το «σχετικό κόστος» που έχουν για μας τα σφάλματα τύπου I και τύπου II. Στο παράδειγμά μας, αν ο ερευνητής απορρίψει τελικά την υπόθεση H_0 ενώ αυτή είναι αληθινή (σφάλμα τύπου I), κινδυνεύει να ξανακάνει χωρίς λόγο τη δειγματοληψία. Αν, αντίστροφα, δεχτεί την υπόθεση της αντιπροσωπευτικότητας του δείγματος ενώ αυτό είναι μεροληπτικό, κινδυνεύει να δώσει λάθος αποτελέσματα.

3) *Υπολογισμός του στατιστικού τεστ.* Είδαμε στη διατύπωση των υποθέσεων ότι οι μέσοι όροι των τυχαίων δειγμάτων μεγέθους N ακολουθούν την κανονική κατανομή με μέσο όρο μ και τυπική απόκλιση $\frac{\sigma}{\sqrt{N}}$. Δηλαδή, στο συγκεκριμένο παράδειγμα, όπου $\mu = 150.000$, $\sigma = 50.000$ και $N = 100$, η δειγματοληπτική κατανομή των μέσων των τυ-

χαίων δειγμάτων θα είναι κανονική με μέσο όρο $\mu = 150.000$ και τυπική απόκλιση $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} = 5.000$. Επομένως, η μεταβλητή:

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - 150.000}{5.000},$$

όπου $\bar{x}$ είναι η μεταβλητή που εκφράζει τους μέσους όρους των τυχαίων δειγμάτων, ακολουθεί την τυποποιημένη κανονική κατανομή. Στο δείγμα που πήραμε το $\bar{x}$ είχε την τιμή 165.000, κι επομένως το Z έχει την τιμή:

$$Z = \frac{165.000 - 150.000}{5.000} = 3.$$

Αυτό σημαίνει ότι ο μέσος όρος του δείγματος που πήραμε αποκλίνει από τον μέσο όρο της δειγματοληπτικής κατανομής 3 μονάδες τυπικής απόκλισης.

4) *Στατιστική απόφαση.* Από τον πίνακα της αθροιστικής συνάρτησης της τυποποιημένης κανονικής κατανομής βρίσκουμε ότι, με πιθανότητα 0.95 (δηλαδή για το επίπεδο σημαντικότητας 0.05 που έχουμε επιλέξει), ο $\bar{x}$ είναι μικρότερος από $\mu + 1{,}65\sigma_{\bar{x}}$. Επομένως, αφού στο παράδειγμά μας ο $\bar{x}$ είναι ίσος με $\mu + 3\sigma_{\bar{x}}$, πρέπει ο ερευνητής να απορρίψει, στο επίπεδο σημαντικότητας 0.05, την υπόθεση (H_0), δηλαδή την υπόθεση της αντιπροσωπευτικότητας του δείγματος.

Στην πραγματικότητα, γνωρίζοντας ακριβώς την τιμή του Z, μπορούμε να πούμε ότι η πιθανότητα να έχουμε ένα Z μεγαλύτερο ή ίσο απ' αυτό που βρήκαμε είναι 0.00135. Η πιθανότητα αυτή μάς καθορίζει το ακριβές επίπεδο σημαντικότητας για το οποίο μπορούμε ν' απορρίψουμε την υπόθεση H_0, και το οποίο συχνά —όπως στο παράδειγμά μας— είναι μικρότερο απ' αυτό που έχουμε εκ των προτέρων επιλέξει.