

Μάθημα 7

§4.2 Ταξινόηση διακριτών αριθμητικών δεδομένων σε κατανομές συχνοτήτων

Η διαδικασία ταξινόμησης των τιμών μιας διακριτής αριθμητικής μεταβλητής σε πίνακα συχνοτήτων είναι η ίδια ουσιαστικά με εκείνη της επεξεργασίας των ονομαστικών δεδομένων. Αντί των συμβολικών τιμών των ονομαστικών δεδομένων καταμετρούμε τις αριθμητικές τιμές της μεταβλητής, καταρτίζοντας έναν πίνακα απόλυτων και σχετικών συχνοτήτων. Δηλαδή κατασκευάζουμε ασυνεχείς (διακριτές) απειρισμένες κατανομές (discrete distributions).

Για παράδειγμα, για τους σκοπούς μιας κοινωνιολογικής έρευνας μελετήθηκε το μέγεθος του νοικοκυριού ως "ο αριθμός των μελών που κατοικούν σ' αυτά". Οι 55 ερωτώμενοι αποτέλεσαν το επιλεγέν δείγμα και έδωσαν τα ακόλουθες απαντήσεις:

ΠΙΝΑΚΑΣ 4.12

Αριθμός μελών σε 55 νοικοκυριά: Αρχικά δεδομένα

2	4	4	1	1	4	1	1	1	2	2
2	3	1	3	2	3	2	2	4	4	1
3	2	1	2	2	6	4	2	2	2	3
2	2	1	2	3	1	1	2	3	4	2
1	2	3	4	1	2	3	2	1	1	4

Στον Πίνακα 4.13 διατάσσουμε τα παραπάνω δεδομένα κατά αύξουσα τάξη μεγέθους και τα στοιχεία του Πίνακα 4.12 συνοψίζονται στον Πίνακα Συχνοτήτων 4.14.

Στην πρώτη στήλη αναγράφονται οι τιμές της εξεταζόμενης μεταβλητής (x_i) και οι διασπόμενες τάξεις, οι υπόλοιπες τέσσερις στήλες ερμηνεύονται ανάλογα με τα όσα είχαμε αναφέρει για τα ονομαστικά δεδομένα. Από τον

Πίνακα 4.14, προκύπτει ότι λίγο περισσότερα από τα μισά νοικοκυριά (50,9%) είναι διμερή, ενώ τα νοικοκυριά στα οποία διαμένουν μέχρι και 3 παιδιά αντιπροσωπεύουν το 87,3% του συνόλου.

Ακόρα και στην περίπτωση της παρουσίας διακριτών δεδομένων συνηθίζεται μερικές φορές η παρουσία τους κατά τάξεις. Η φύση και ο αριθμός των παρατηρήσεων των στοιχείων του Παραδείγματος του Πίνακα 4.12 δεν δικαιολογούν την κατάρτιση περισσότερων από τρεις τάξεις:

	f_i	$f_i \%$
από 1 έως και 2 μέλη	34	61,82
από 3 έως και 4 μέλη	18	32,73
από 5 έως και 6 μέλη	3	5,45
Σύνολο Παρατηρήσεων	55	100,00

Στις κατανομές διακριτών δεδομένων, το πλάτος των διαστημάτων τάξεων προκύπτει ως διαφορά μεταξύ των ανώτερων ορίων δύο διαδοχικών τάξεων δηλ. $\delta i = L_{i+1} - L_i$ και όχι μεταξύ του ανώτερου και του κατώτερου ορίου της ίδιας τάξης όπως συμβαίνει στις κατανομές συνεχούς μεταβλητής.

§4.3 Γραφική απεικόνιση κατανομών συχνότητων

Η γραφική απεικόνιση μίας κατανομής συχνότητων συνεχών δεδομένων μπορεί να γίνει με τρεις τρόπους:

- (α) Με Ιστόγραμμα συχνότητων (frequency histogram)
- (β) Με Πολύγωνο συχνότητων (frequency polygon)
- (γ) Με Καμπύλη συχνότητων (frequency curve)

Τα γραφήματα αυτά παρουσιάζουν σχηματικά τη δομή του εξεταζόμενου πληθυσμού ή δείγματος.

Ιστογράμματα (Histograms): Είναι ο σημαντικότερος τρόπος γραφικής απεικόνισης μίας συνεχούς κατανομής συχνοτήτων. Με το ιστογράμμα μπορούμε με μία ματιά να ελέγχουμε το βαθμό συχνέντρωσης των δεδομένων σε κάποια ή κάποιες από τις τάξεις καθώς και το ποσοστό των παρατηρήσεων που ερπίζει μεταξύ δύο τιμών της μεταβλητής.

Η διαφορά ανάμεσα στα ραβδογράμματα και στα ιστογράμματα έγκειται στο ότι τα ραβδογράμματα χρησιμοποιούνται για την απεικόνιση ποιοτικών μεταβλητών ενώ τα ιστογράμματα χρησιμοποιούνται για την απεικόνιση ποσοτικών μεταβλητών. Τα ορθογώνια παραλληλόγραμμα των ραβδογραμμάτων (αιίδες) σχεδιάζονται ώστε να απέχουν μεταξύ τους. Στα ιστογράμματα τα ορθογώνια (ιστοί) είναι εφάπτομενα μεταξύ τους.

Για να κατασκευάσουμε ένα ιστογράμμα με τάξεις ίδιου πλάτους:

- Θεωρούμε το πρώτο τεταρτημόριο του συστήματος των ορθογωνίων συντεταγμένων.
- Κατά μήκος του άξονα x απεικονίζουμε τα διαστήματα τάξεων της κατανομής με διαδοχικά ευθύγραμμα τμήματα συνήθως ίδιου μήκους.
- Κατά μήκος του άξονα y απεικονίζουμε την μέγιστη μέτρηση των συχνοτήτων (απόλυτων ή σχετικών) με σημείο εκκίνησης το 0.
- Με βάση τα ευθύγραμμα τμήματα που ορίσαμε στον οριζόντιο άξονα σχεδιάζουμε τόσα ορθογώνια παραλληλόγραμμα (ιστούς) όσες και οι τάξεις της κατανομής με ύψος που αναλογεί στην συχνότητα της αντίστοιχης

τάξης. Τα ορθογώνια που σχηματίζονται έχουν το ίδιο πλάτος για όλες τις τάξεις της κατανομής, το τέλος της βάσης κάθε ορθογωνίου που αντιστοιχεί στο ανώτερο όριο τάξης ταυτίζεται με την αρχή του αμέσως επόμενου δηλαδή αντιστοιχεί στο κατώτερο όριο της επόμενης τάξης.

Στον πίνακα 4.3 παρουσιάσαμε την κατανομή 30 ατόμων ανάλογα με το βάρος τους σε πέντε τάξεις πλάτους 5 κιλών. Το ιστόγραμμα απόλυτων συχνοτήτων παρουσιάζεται στο Σχήμα 4.1α. Κάθε ένα από τα πέντε εφάπτομενα παραλληλόγραμμα αντιστοιχεί σε κάθε ένα από τα ισόριθρα διαστήματα τάξεων. Το ύψος των παραλληλογράφων απεικονίζει αναλογικά τις αντίστοιχες συχνότητες και η βάση τους το πλάτος των τάξεων. Το εμβαδόν όλης της περιοχής που περιέχεται μέσα στο σύνολο των ορθογωνίων αντιστοιχεί στο σύνολο ($n=30$) δηλαδή στο 100% των παρατηρήσεων. Το εμβαδόν κάθε επί μέρους ορθογωνίου αναλογεί προς την απόλυτη συχνότητα της αντίστοιχης τάξης. Ο λόγος του εμβαδού κάθε παραλληλογράφου προς το συνολικό εμβαδόν του σχήματος είναι ίσο με $\frac{f_i}{n}$ δηλαδή με τη σχετική συχνότητα. Για το παράδειγμά μας, μπορούμε να πούμε ότι $6+12=18$ από τα 30 εξεταζόμενα άτομα (δηλαδή αναλογία 0.6) έχουν βάρος 70 έως 80 κιλά (Σχήμα 4.1(β)).

Πολύγωνο Συχνοτήτων: Παρουσιάζει με πιο ενδιαφέρον τρόπο απ' ό,τι το ιστόγραμμα τη μορφή των στοιχείων.

Χρησιμοποιείται για τη διαγραμματική απεικόνιση συνεχών δεδομένων.

- Για την κατασκευή του καταρτίζουμε το ιστόγραμμα των απλών (απόλυτων ή σχετικών) συχνοτήτων της κατανομής.
- Εντοπίζουμε τα κέντρα των άνω πλευρών των ορθογωνίων (ιστίων) και τα συνδέουμε μεταξύ τους με ευθύγραμμα τμήματα.
- Η πολυγωνική γραμμή που σχηματίζεται γ' αυτόν τον τρόπο ξεκινά από το μέσο του διαστήματος που βρίσκεται πριν από το πρώτο ορθογώνιο και καταλήγει στο μέσο του διαστήματος που βρίσκεται μετά το τελευταίο ορθογώνιο ώστε να "υψώσει" η πολυγωνική γραμμή και να δώσει μοναδιαίο εμβαδόν της περιχλειόμενης επιφάνειας.

Το πολύγωνο των απόλυτων συχνοτήτων της κατανομής των βαρών των 30 ατόμων του πίνακα 4.3 παρουσιάζεται στο Σχήμα 4.2α. Η παρουσίαση του συχνοπολύγωνου δεν συνοδεύεται απαραίτητα από το αντίστοιχο ιστόγραμμα, όπως φαίνεται στο Σχήμα 4.2(β).

Καρπύλη Συχνοτήτων: Είναι προφανές ότι αν αυξηθεί πάρα πολύ το πλήθος των τάξεων και συνεπώς μειωθεί το πλάτος των διαστημάτων τάξεων η πολυγωνική γραμμή που σχηματίζεται από τη σύνδεση των κέντρων των κορυφών των ορθογωνίων παραλληλογράμμων του αντίστοιχου ιστογράμματος συχνοτήτων προσεγγίζει μία καρπύλη που καλείται καρπύλη συχνοτήτων. Είναι μία γραφική απεικόνιση απείρων σημείων των οποίων οι

τετρημένες αντιστοιχούν στις άπειρες τιμές της συνε-
χούς μεταβλητής. Το σχήμα 4.4 εμφανίζει την καρπύλη
κατανομής των χάρων του 1994 κατά ηλικία της νέας.
Χρησιμοποιήθηκαν επίσης δηλωτικές καταγραφές
που δημοσιεύονται από την ΕΣΥΕ. Η θεωρητική μελέ-
τη των καρπύλων συχνοτήτων μας οδηγεί στην μαθη-
ματική διατύπωση σχέσεων που εκφράζουν τις τιμές
της μεταβλητής και τις συχνότητες εμφάνισής τους
στον πληθυσμό που μελετάμε.

Γενικά διακρίνουμε τέσσερις κατηγορίες κατανομών
που λαμβάνουν το όνομά τους από το σχήμα που παρου-
σιάζουν: (α) Μονομόρφες κατανομές (Unimodal distribu-
tions), (β) κατανομές σχήματος U (U-shaped distributions)
(γ) κατανομές σχήματος J (J-shaped distributions) και
(δ) σύνθετες κατανομές.

Το εμβαδόν της επιφάνειας που περιλαμβάνεται από τις
καρπύλες και τον οριζόντιο άξονα αντιστοιχεί στο σύνολο
των συχνοτήτων. Οποιοδήποτε τμήμα του εμβαδού δίνει
την αναλογία των δεδομένων που περιέχεται μεταξύ
δύο τιμών της μεταβλητής. Η κατανομή του 4.5(α)
είναι συμμετρική (symmetrical distribution) ως προς
τον κάθετο άξονα που διέρχεται από την κορυφή
της καρπύλης και καλείται κωδωνοειδής καρπύ-
λη (bell shaped distribution). Ο άξονας συμμετρίας
χωρίζει την επιφάνεια που ορίζεται από την καρπύ-
λη της κατανομής σε δύο ίσα τμήματα καθένα των
οποίων περιλαμβάνει αριθμώς το 50% των παρατη-
ρήσεων. Τα δύο άκρα (αριστερά και δεξιά) της κα-
ρπύλης καλούνται ουρές (tails) και τείνουν ασυμπτωτικά

προς τον άξονα των x' . Οι καρπύρες που παρουσιάζονται στα σχήματα 4.5β και 4.5γ είναι επίσης μονοκύρμες δεν είναι όμως συρρετικές αφού αριστερά και δεξιά από το μέγιστο σημείο τους (κορυφή της αντίστοιχης καρπύρας) οι συχνότητες φθίνουν ασύμμετρα. Η καρπύρη του σχήματος 4.5β αναφέρεται σε κατανομή αρνητικής ή αριστερής ασυρρετίας (distribution skewed to the left) λόγω ραυρότερης αριστερής ουράς ενώ η καρπύρη του σχήματος 4.5γ αντιστοιχεί σε κατανομή θετικής ή δεξιάς ασυρρετίας (distribution skewed to the right) λόγω ραυρότερης δεξιάς ουράς της κατανομής.

Κατανομές σχήματος U ή αντιστρόφου U (Σχήμα 4.5δ και 4.5ε) παρατηρούνται σε οικονομικά μεγέθη αλλά γενικά δεν συναντώνται στις εφαρμογές. Οι κατανομές σχήματος J και αντίστροφου J (Σχήμα 4.5στ και 4.5ζ) δεν είναι σνήθεις. Οι κατανομές της τελευταίας κατηγορίας περιγράφουν σύνθετα και πολύπλοκα φαινόμενα, η μελέτη των οποίων δεν θα μας απασχολήσει. Συνήθως, το ενδιαφέρον των ερευνητών επικεντρώνεται στις μονοκύρμες κατανομές.

Διατάσσουμε τα ανωτέρω δεδομένα κατά αύξουσα τάξη μεγέθους:

(4.1.8)

ΠΙΝΑΚΑΣ 4.13

Αριθμός μελών σε 55 νοικοκυριά: διατεταγμένα δεδομένα

1	1	1	1	1	1	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
2	3	3	3	3	3	3	3	3	3	3
3	3	3	3	4	4	4	4	5	5	6

Πηγή: Στοιχεία Πίνακα 4.12

Τα στοιχεία του Πίνακα 4.12 συνοψίζονται στον πιο κάτω πίνακα συχνοτήτων:

ΠΙΝΑΚΑΣ 4.14

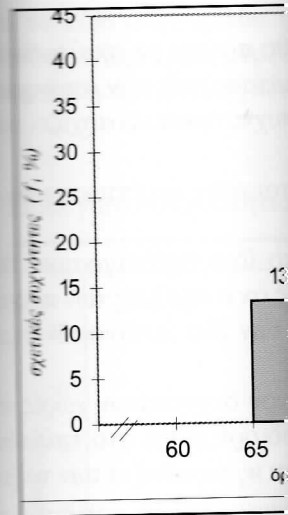
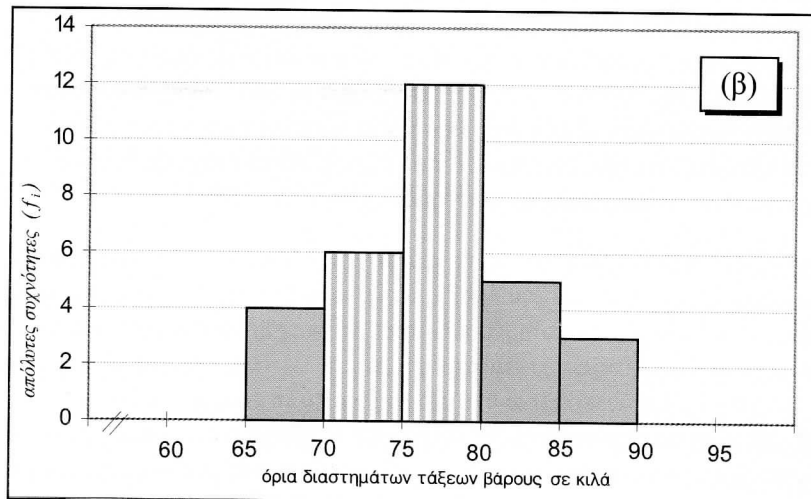
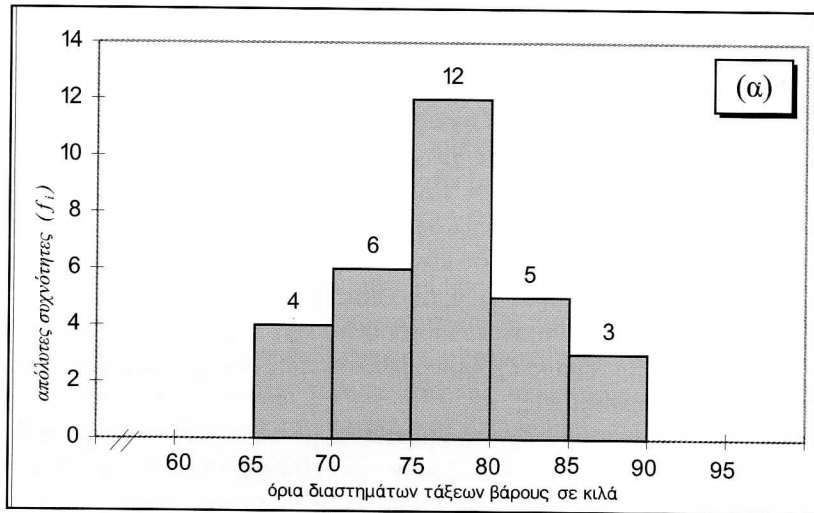
Κατανομή 55 νοικοκυριών κατά μέγεθος

Αριθμός Μελών	Αριθμός Νοικοκυριών	Ποσοστό (%) Νοικοκυριών	Αθροιστική Συχνότητα	Αθροιστική Συχνότητα (%)
(X_i)	f_i	$f_i(\%)$	F_i	$F_i(\%)$
1	6	10,91	6	10,91
2	28	50,91	34	61,82
3	14	25,45	48	87,27
4	4	7,27	52	94,55
5	2	3,64	54	98,18
6	1	1,82	1	100,00
Σύνολο	55	100,0		

Πηγή: Στοιχεία Πίνακα 4.13

ΣΧΗΜΑ 4.1

Ιστόγραμμα συχνοτήτων της κατανομής βάρους του Πίνακα 4.3



4.3.2 Πολύγωνο συχνοτήτων

Το πολύγωνο συχνοτήτων αποτελεί το ιστόγραμμα της διαγραμματικής απεικόνισης για τη ταυτόχρονη απεικόνιση της κατανομής της εξεταζόμενης κατανομής.

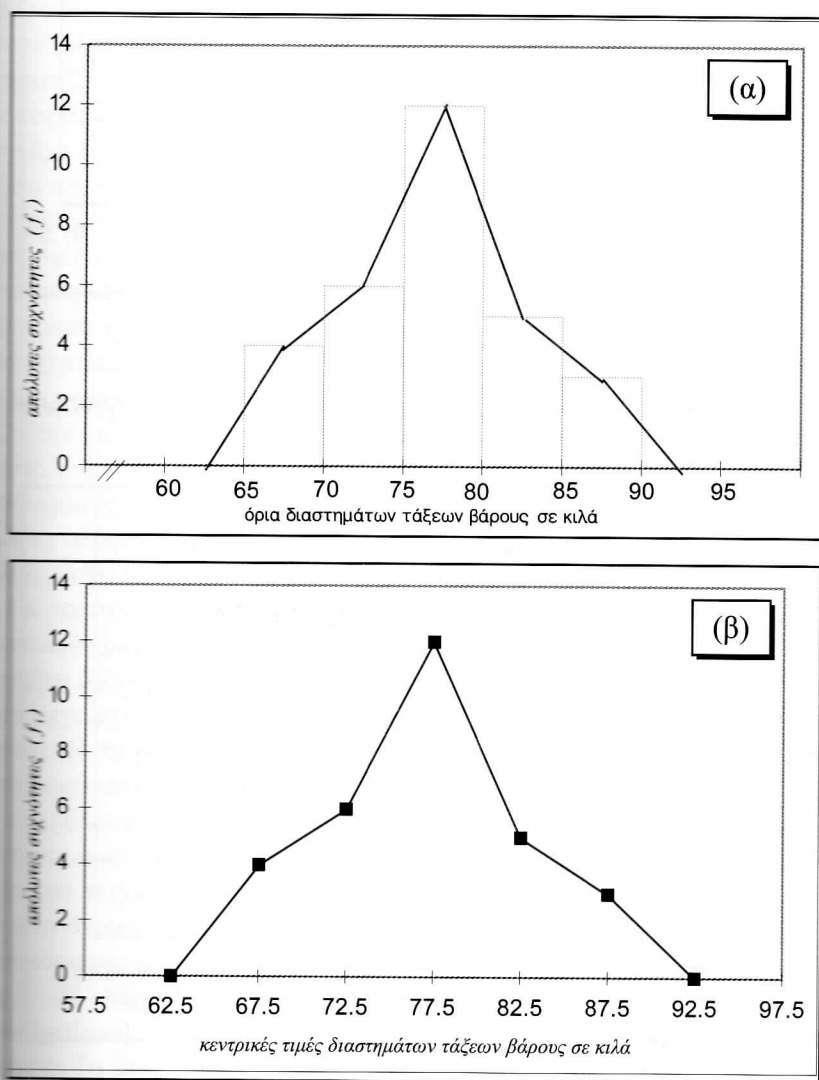
Τεχνική
συχνότητας συνεχούς

- Καταρτίζουμε το ιστόγραμμα συχνοτήτων της κατανομής πάνω.
- Εντοπίζουμε τα κέντρα των βαρών και τα συνδέουμε με ευθείες γραμμές.
- Η πολυγωνική γραμμή που σχηματίζεται από το μέσον του διαστήματος και καταλήγει στο τελευταίο ορθογώνιο (κλείστρο) ώστε να "κλείσει" ο εμβαδόν της περιμέτρου.

ενώ το 1988 στην ομάδα ηλικιών 20-25 ετών. Παρατηρείται επίσης δια-
χρονικά αύξηση των αναλογιών στις μικρές ηλικίες (μέχρι τα 25 έτη) και
μία μείωση των αναλογιών από το σημείο αυτό και άνω.

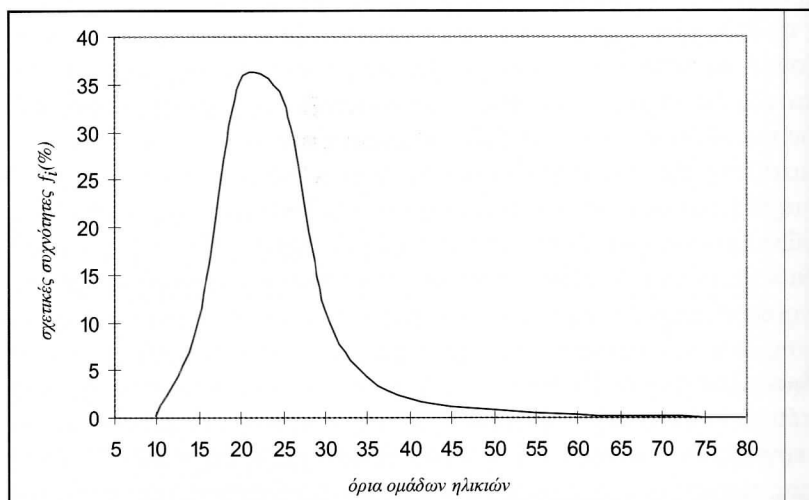
ΣΧΗΜΑ 4.2

Πολύγωνο συχνοτήτων της κατανομής βάρους του Πίνακα 4.3



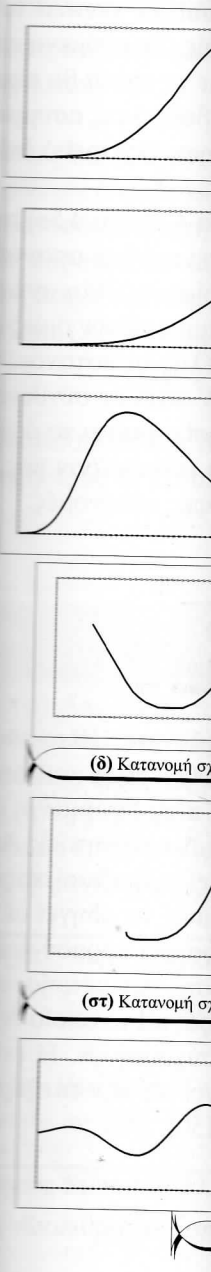
ΣΧΗΜΑ 4.4

Καμπύλη συχνοτήτων της κατανομής των γάμων κατά ηλικία της
νύφης: Ελλάδα 1994



Ιδιαίτερης σημασίας είναι οι **μονοκόρυφες κατανομές**. Το εμβαδόν της επιφάνειας που περικλείεται από τις καμπύλες και τον οριζόντιο άξονα αντιστοιχεί στο σύνολο των συχνοτήτων. Οποιοδήποτε τμήμα αυτού του εμβαδού μπορεί να εκτιμηθεί και να δώσει την αναλογία των δεδομένων που περιλαμβάνεται μεταξύ δύο τιμών της μεταβλητής. Η πρώτη από τις κατανομές που παρουσιάζονται (Σχήμα 4.5α) είναι **συμμετρική (symmetrical distribution)** ως προς τον κάθετο άξονα που διέρχεται από την κορυφή της καμπύλης και καλείται **κωδωνοειδής κατανομή (bell shaped distribution)**. Ο άξονας συμμετρίας χωρίζει την επιφάνεια που ορίζεται από την καμπύλη της κατανομής σε δύο ίσα τμήματα καθένα των οποίων περιλαμβάνει ακριβώς το 50% των παρατηρήσεων. Τα δύο άκρα (αριστερά και δεξιά) της καμπύλης λέγονται **ουρές (tails)** και τείνουν ασυμπτωτικά προς τον άξονα των XX' . Η ονομαζόμενη **κανονική κατανομή (normal distribution)** καθώς και η **τυποποιημένη κανονική κατανομή (standard normal distribution)** έχουν αυτή την συμμετρική κωδωνοειδή μορφή, λόγω δε των ιδιοτήτων που τις χαρακτηρίζει χρησιμοποιούνται ευρέως στην θεωρία και στις εφαρμογές της επαγωγικής στατιστικής. Οι καμπύλες που παρουσιάζονται στα Σχήματα 4.5β και 4.5γ είναι επίσης μονοκόρυφες δεν είναι όμως συμμετρικές αφού αριστερά και δεξιά από το μέγιστο σημείο τους (κορυφή της αντίστοιχης καμπύλης) οι συχνότητες φθίνουν ασύμμε-

Σχηματική μ



ΣΧΗΜΑ 4.5

Σχηματική μορφή θεωρητικών συνεχών κατανομών.

