

Μάθημα 6

(β) Ραβδογράμματα ή Αιιδωτά διαγράμματα: Αποτελούν από αλλα καθιερωμένο τρόπο γραφικής απεικόνισης των απόλυτων και σχετικών συχνότητων της κατανομής ρίας ποιοτικής μεταβλητής.

Θεωρούμε το σύστημα των ορθογωνίων συντεταγμένων. Κατά μήκος του άξονα Y απεικονίζουμε την κλίμακα μέτρησης των συχνότητων (απόλυτων ή σχετικών) ξεκινώντας από το 0. Κατά μήκος του οριζόντιου άξονα X απεικονίζουμε τις κατηγορίες της μεταβλητής με ευθύγραμμα τμήματα ορθογωνίου μήκους που πρέπει να είναι ίσα μεταξύ τους, να μην εφάπτονται και να ισαπέχουν. Με βάση τα ευθύγραμμα τμήματα του οριζόντιου άξονα σχεδιάζουμε τόσα ορθογώνια παραλληλόγραμμα (αιίδες) όσα οι κατηγορίες της ποιοτικής μεταβλητής με ύψος που αναλογεί στη συχνότητα της αντίστοιχης κατηγορίας.

Στο Σχήμα 3.2α γίνεται η απεικόνιση των απόλυτων συχνότητων και στο Σχήμα 3.2β των σχετικών (%). Είναι στην κρίση του ερευνητή να εμφανίσει ή να μην εμφανίσει τις συχνότητες της κατανομής στις κορυφές των ορθογωνίων παραλληλόγραμμων. Στα Σχήματα 3.2γ και 3.2δ έχουν προστεθεί πλέγματα δηλαδή οριζόντιες ή/και κάθετες γραμμές οι οποίες

διευκολύνουν τον αναγνώστη να "διαβάσει" το σχήμα. -30-
Στο Σχήμα 3.2δ οι συχνότητες έχουν μετρηθεί στον
άξονα των Χ με συνέπεια τα ραβδογράμματα να
σχεδιάζονται οριζόντια.

Στον Πίνακα 3.4 παρουσιάζονται ο αριθμός των απασχο-
λούμενων κατά φύλο και κατά επίπεδο αστιότητας
Ένας τέτοιος πίνακας καλείται πίνακας διπλής εισόδου
διότι παρέχει πληροφορίες για δύο πληθυσμιακά
χαρακτηριστικά ταυτόχρονα. Σε αυτές τις περιπτώ-
σεις τα ραβδογράμματα απεικονίζουν δύο πληροφορίες
(φύλο και αστιότητα) και παρέχουν τη δυνατότητα
διπλής ~~απτιικής~~ σύγκρισης της απασχόλησης των ανδρών
και των γυναικών κατά επίπεδο αστιότητας. Επειδή
εμφανίζει ανά κατηγορία εφαιπόμενες αίδες καλείται
συζυγές ή ενοποιημένο ραβδόγραμμα. Τα ίδια στοιχεία
θα μπορούσαν να παρουσιαστούν γραφικά με ένα σω-
ρευτικό ή πυρροειδές ραβδόγραμμα. Στο Σχήμα 3.4β
η κορυφή κάθε πυρροειδούς αίδας αντιστοιχεί στο
συνολικό αριθμό των απασχολούμενων στις αστικές,
ημιαστικές και αγροτικές περιοχές. Στα Σχήματα
3.4γ και 3.4δ, στον οριζόντιο άξονα μετράμε το φύλο
των απασχολούμενων και στον κάθετο άξονα τον
αριθμό των απασχολούμενων των αστικών, ημιαστι-
κών και αγροτικών περιοχών. Στο Σχήμα 3.4δ, στον
κάθετο άξονα έχουμε μετρήσει τις σχετικές συχνό-
τητες της κατανομής.

ΑΡΙΘΜΗΤΙΚΕΣ ΚΑΙ ΓΡΑΦΙΚΕΣ ΤΕΧΝΙΚΕΣ ΠΕΡΙΓΡΑΦΗΣ ΑΡΙΘΜΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

§4.1 Σύρτυξη συνεχών δεδομένων σε κατανομές συχνότητας

Η δυσκολία στο χειρισμό πολυπληθών παρατηρήσεων πολλές φορές στην σύρπυνση της παρεχόμενης πληροφορίας. Γι' αυτό χρησιμοποιείται η κατάταξη των συνεχών δεδομένων σε κλάσεις (τάξεις, ομάδες) και η παρουσίασή τους σε πίνακες κατανομής συχνότητας παρόμοιους με αυτούς των ονομαστικών μεταβλητών.

Τα στατιστικά στοιχεία συλλέγονται κατά τυχαίο τρόπο χωρίς καμία λογική σειρά.

Για παράδειγμα τα πρωτογενή (ανατέρχαστα) δεδομένα του βάρους 50 ατόμων φαίνονται στον Πίνακα 4.1.

Για να διευκολύνουμε την ταξινόμηση των αρχικών δεδομένων σε πίνακα κατά συχνότητα εμφάνισης

διατάσσουμε τα αρχικά μας δεδομένα κατά αύξουσα τάξη μεγέθους (Πίνακας 4.2). Η ελάχιστη τιμή της μεταβλητής είναι 65,1 κιλά και η μεγαλύτερη είναι 89,9 κιλά. Εύρος (range) συμβολίζουμε με R και δίνει τη διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής της μεταβλητής. Η μέγιστη τιμή συμβολίζεται με $x_{\max}$ και η ελάχιστη με $x_{\min}$, δηλαδή

$$R = x_{\max} - x_{\min} = 89,9 - 65,1 = 24,8 \text{ κιλά.}$$

Για να κατασκευάσουμε μία κατανομή των δεδομένων κατά συχνότητες θα πρέπει να υποδιαιρέσουμε το εύρος (R) των τιμών της μεταβλητής σε τάξεις ή διαστήματα τάξεων (class intervals) ώστε να αποφευχθούν επικυψώσεις ή κενές περιοχές τιμών. Η καταρέτρωση των τιμών της μεταβλητής που ερπύπτουν σε κάθε τάξη ολογηγρώνει την οραδοποίηση των δεδομένων. Στις συνεχείς κατανομές οι συχνότητες καδούνται ταξιμένες συχνότητες. Το πλήθος των τάξεων συμβολίζεται με K και εξαρτάται από τη φύση των δεδομένων και από το εύρος των τιμών τους. Συνήθως ένας αριθμός τάξεων μεταξύ 5 και 20 θεωρείται ικανοποιητικός. Πρακτικά, η χρήση κατανομών συχνοτήτων με ίσο πλάτος διευκολύνει την ανάγνωση των στοιχείων. Το πλάτος των διαστημάτων (τάξεων) προσδιορίζονται από τον τύπο: $\text{πλάτος διαστημάτων τάξεων} = \delta = \frac{R}{K}$, όπου K ο επιθυμητός αριθμός των τάξεων της κατανομής. Ο εμπειρικός τύπος του Sturges παρέχει κατά προσέγγιση το πλάτος των διαστημάτων τάξεων:

$$\delta = \frac{R}{1 + 3,322 \log n}, \text{ όπου } n \text{ ο αριθμός}$$

των παρατηρήσεων. Με βάση τα διατεταγμένα στοιχεία του Πίνακα 4.2 υπολογίζουμε το δ ως εξής:

$$\delta = \frac{89,9 - 65,1}{5} = \frac{24,8}{5} = 4,96 \approx 5.$$

Η πρώτη τιμή της μεταβλητής (65,1 κιλά) θα περιέχεται στην πρώτη τάξη η κατανομή θα ξεκινά από την στοχαστοποιημένη τιμή των 65 κιλών (Πίνακας 4.3).

Οι κατώτερες και ανώτερες τιμές που προσδιορίζουν τα διαστήματα τάξεων καλούνται όρια τάξεων. Π.χ. στην τρίτη τάξη, το κατώτερο όριο είναι 75 κιλά και το ανώτερο είναι τα 80 κιλά. Τα κατώτερα και τα ανώτερα όρια των διαστημάτων τάξεων συμβολίζονται με L_i και U_i αντίστοιχα. Η διαφορά των δύο ορίων προσδιορίζει το πλάτος των διαστημάτων $\delta_i, i=1, \dots, k$ δηλαδή $\delta_i = U_i - L_i$.

Το ημιάθροισμά τους καλείται κεντρική τιμή διαστημάτων τάξεων και συμβολίζεται με

$$W_i = \frac{L_i + U_i}{2}$$

ιάνουμε είναι ότι οι συχνότητες κατανέμονται ομοιόμορφα μέσα στα διαστήματα τάξεων και συνεπώς συχμεντρώνονται γύρω από την κεντρική τιμή της αντίστοιχης τάξης. Για παράδειγμα, από τα 30 άτομα που εξετάστηκαν, τα 4 (ή το 13,3% του συνόλου βρέθηκαν με βάρος 65 έως 70 κιλά. Η άγνοια των αμειβών βαρών μας υποχρεώνει να ιάνουμε την παραδοχή ότι καθένα των τεσσάρων ατόμων έχει βάρος ίσο με $\frac{(65+70)}{2} = 67,5$ κιλά.

Το άθροισμα των συχνοτήτων όλων των τάξεων είναι
ίσο με το σύνολο των διαθέσιμων παρατηρήσεων και το
αντίστοιχο άθροισμα των σχετιών συχνοτήτων είναι
ίσο με 1 ή με 100 ανάλογα αν οι σχετιές συχνο-
τητες εκφράζονται ανά μονάδα ή ως ποσοστό επί
τους εκατό: $\sum_{i=1}^K f_i = n, \sum_{i=1}^K \frac{f_i}{n} = 1 \text{ ή } \sum_{i=1}^K \frac{f_i}{n} 100 = 100$

όπου K το πλήθος των τάξεων της κατανομής.
Οι απόλυτες αθροιστικές συχνότητες (f_i) υπολογί-
ζονται αθροίζοντας τις απλές συχνότητες f_i από
την πρώτη μέχρι την τελευταία τάξη της κατα-
νομής. Οι σχετιές αθροιστικές συχνότητες εκφρά-
ζονται ως ποσοστό (%) επί του συνολικού αριθμού των
παρατηρήσεων (Πίνακας 4.10).

Δέκα άτομα επί συνόλου 30 (ποσοστό 33,3%) θρέθηκαν
με βάρος μικρότερο των 75 κιλών ενώ το 90% των εξε-
ταζόμενων (ή 27 άτομα) θρέθηκαν να βυγίζουν λιγό-
τερο από 85 κιλά.

Ισχύει ότι:
$$F_i = \sum_{j=1}^i f_j, \quad i=1, 2, \dots, K$$

$$F_1 = f_1$$

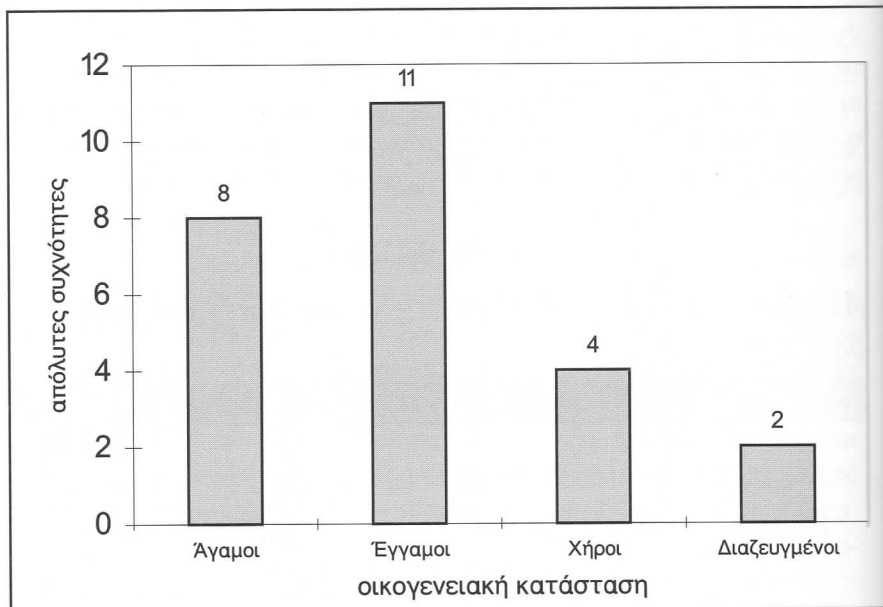
$$F_{i+1} - F_i = f_i \Leftrightarrow F_{i+1} = F_i + f_i, \quad i=1, \dots, K, \text{ όπου}$$

K το πλήθος των τάξεων της κατανομής.

Συχνές είναι οι περιπτώσεις δεδομένων στα οποία παρου-
σιάζεται μικρός αριθμός παρατηρήσεων με ασυνήθιστα υψη-
λές ή χαμηλές τιμές σε σχέση με τις υπόλοιπες. Αυτές οι
"ιδιαίτερες τιμές" επιμάρτησαν στη βιβλιογραφία με τον
χαρακτηρισμό "έκτροπες τιμές" ή "έκτροπες παρατηρήσεις".

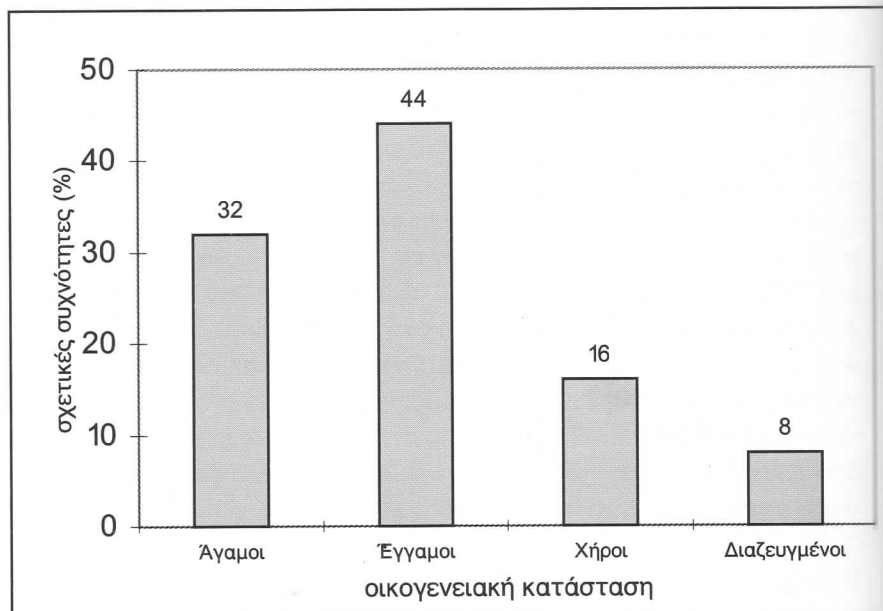
ΣΧΗΜΑ 3.2α

Ραβδόγραμμα των στοιχείων του Πίνακα 3.1



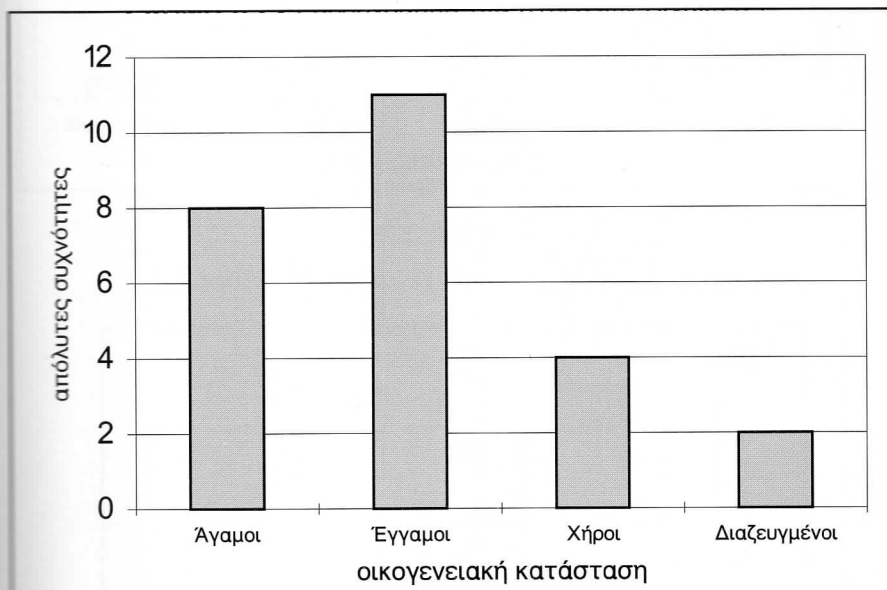
ΣΧΗΜΑ 3.2β

Ραβδόγραμμα των στοιχείων του Πίνακα 3.1



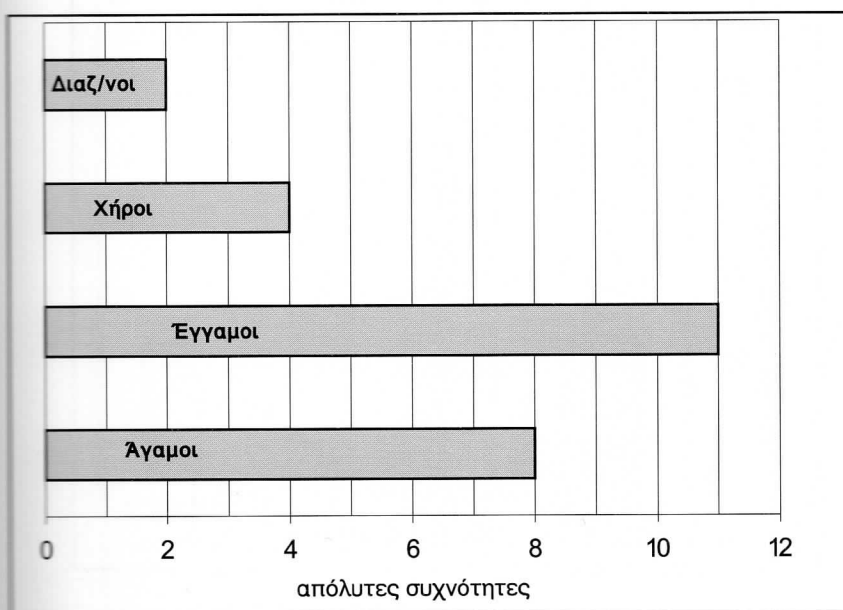
ΣΧΗΜΑ 3.2γ

Ραβδόγραμμα των στοιχείων του Πίνακα 3.1



ΣΧΗΜΑ 3.2δ

Ραβδόγραμμα των στοιχείων του Πίνακα 3.1



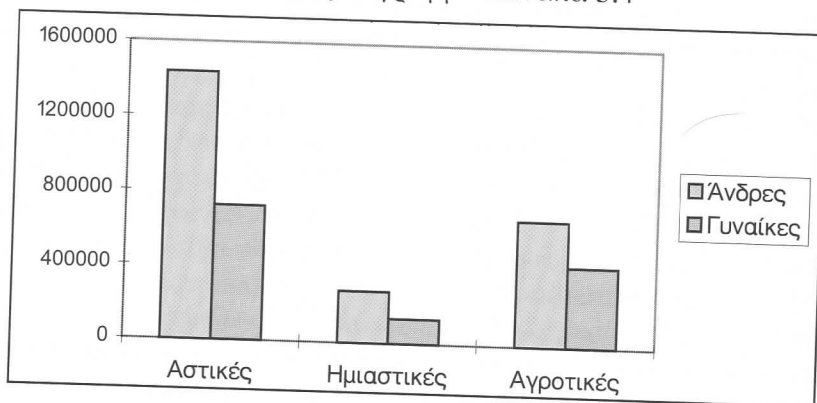
ΠΙΝΑΚΑΣ 3.4

Απασχολούμενοι κατά φύλο και περιοχή συνήθους διαμονής: 1989

Περιοχές	Αριθμοί		Ποσοστά	
	Άνδρες	Γυναίκες	Άνδρες	Γυναίκες
Αστικές	1437800	723700	60,27	56,30
Ημιαστικές	275500	132200	11,55	10,28
Αγροτικές	672200	429600	28,18	44,42
Σύνολο	2385500	1285500	100,00	100,00

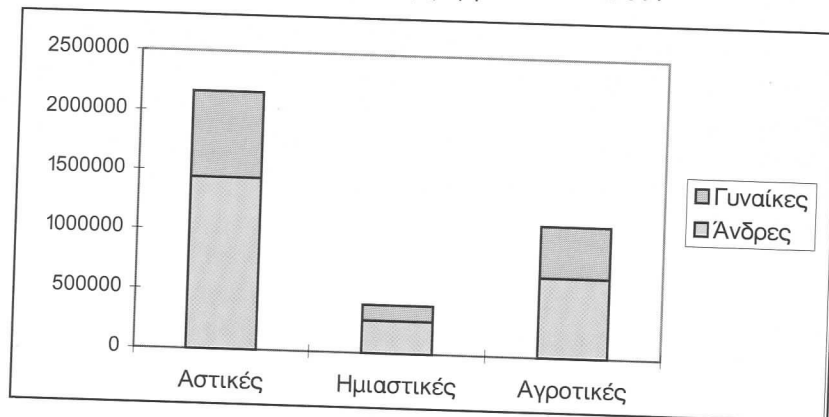
ΣΧΗΜΑ 3.4α

Συζυγές Ραβδόγραμμα Πίνακα 3.4

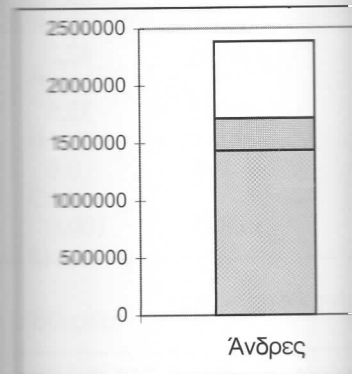


ΣΧΗΜΑ 3.4β

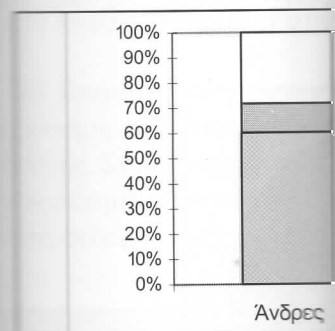
Σωρευτικό Ραβδόγραμμα Πίνακα 3.4



Σωρευτικό Ραβδόγραμμα



Σωρευτικό Ραβδόγραμμα

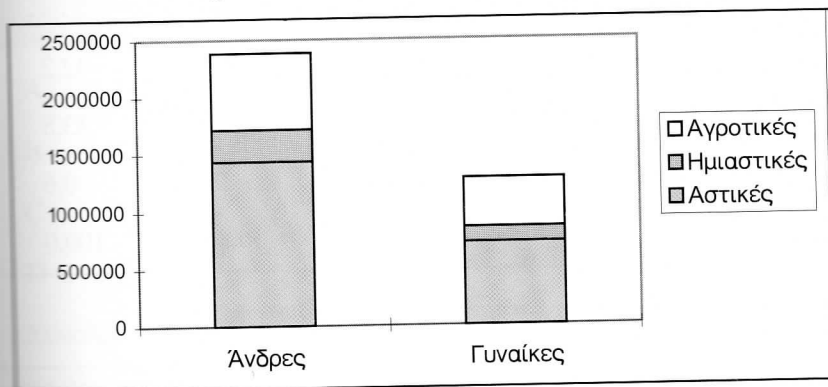


1.5 Ειδικά Θέματα

3.1 Αρχικές φορές συμβαίνει να πολύ μικρή σε σχέση με τις απαιτήσεις παρουσιάζουν τα στοιχεία την κατανομή των υπηρεσιών ανάλογα με το μορφή (π.χ. έτος 1993).

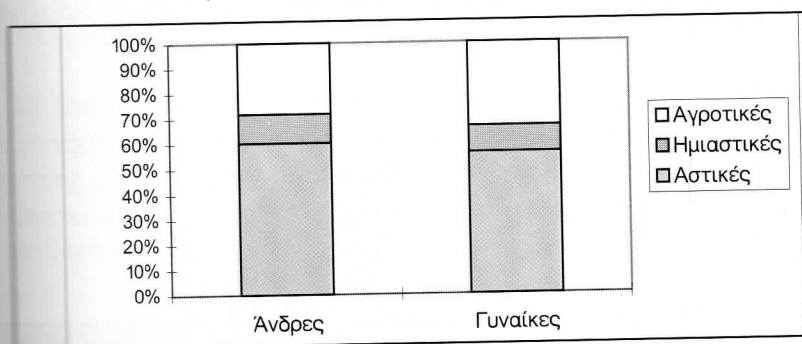
ΣΧΗΜΑ 3.4γ

Σωρευτικό Ραβδόγραμμα Πίνακα 3.4



ΣΧΗΜΑ 3.4δ

Σωρευτικό Ραβδόγραμμα Πίνακα 3.4



3.5 Ειδικά Θέματα

3.1 Αρκετές φορές συμβαίνει η συχνότητα κάποιας κατηγορίας να είναι πολύ μικρή σε σχέση με τις συχνότητες των υπολοίπων. Μία τέτοια κατάσταση παρουσιάζουν τα στοιχεία που δίδονται πιο κάτω, τα οποία παρήχθησαν από την κατανομή των υπηρετούντων νοσοκόμων στα Ελληνικά θεραπευτήρια ανάλογα με το μορφωτικό τους επίπεδο (τα στοιχεία αφορούν το έτος 1993).

μεταβλητής είναι 65,1 κιλά, και η μεγαλύτερη είναι 89,9 κιλά. Συνεπώς η έκταση των τιμών της μεταβλητής είναι $89,9 - 65,1 = 24,8$ κιλά. Ορίζουμε ως **εύρος (range)** και συμβολίζουμε με R τη διαφορά μεταξύ της μέγιστης (x_{max}) και ελάχιστης (x_{min}) τιμής της μεταβλητής. Συνοπτικά:

$$R = x_{max} - x_{min}$$

ΠΙΝΑΚΑΣ 4.1

Βάρος τριάντα ατόμων σε κιλά: αρχικά δεδομένα

79,6	72,3	86,1	75,0	67,3	74,1
76,7	79,9	89,9	76,2	78,6	73,0
85,8	78,0	80,8	81,4	65,1	81,8
76,5	66,8	79,6	77,1	80,1	72,3
78,3	67,9	76,5	73,0	83,9	70,0

ΠΙΝΑΚΑΣ 4.2

Βάρος τριάντα ατόμων σε κιλά: διατεταγμένα δεδομένα

65,1	66,8	67,3	67,9	70,0	72,3
72,3	73,0	73,0	74,1	75,0	76,2
76,5	76,5	76,7	77,1	78,0	78,3
78,6	79,6	79,6	79,6	80,1	80,8
81,4	81,8	83,9	85,8	86,1	89,9

(β) Κατανομή των δεδομένων κατά συχνότητες

Στην περίπτωση των ονομαστικών μεταβλητών, η καταμέτρηση και κα-

πνομή των δεδομένων των ονομαστικών τιμών της περίπτωση, μεταξύ των $(=x_{max})$ κιλών - η εξεταστέος, να λάβει οποιαδήποτε συνεχούς μεταβλητής των εύρους (R) των τιμών (*classes, class intervals*) και πενές περιοχές τιμών. Η της μεταβλητής που επιπλέον ομαδοποίησης (ταξινόμησης) **κατανομές (contingency tables)** **συχνότητες (class frequencies)**

Το πλήθος των τάξεων k . Ο αριθμός των περιπτώσεων από τη φύση των περισσότερες περιπτώσεις που ικανοποιεί τα επιθυμητά κριτήρια και λειτουργία στην χρησιμοποίηση μικρότερων (50% ή και περισσότερων) ή και μηδενικών των δεδομένων, επιπλέον πλέον παρέχεται από τις ομαδοποιήσεις των k αριθμός των υπό μελέτη

Στην πράξη, η χρήση και επιθυμητή διότι διευκολύνει το εύρος τιμών της μεταβλητής **τάξεων (class intervals)**

Πλάτος διαστημάτων

άπου k ο επιθυμητός αριθμός

στρογγυλεύοντας την τιμή του αριθμού που να διευκολύνει

