

§ 5.8 Διάρκεσος (Median) είναι η τιμή που χωρίζει το σύνολο των διατεταγμένων δεδομένων σε δύο ισοπαγή υποσύνολα, ώστε το 50% των παρατηρήσεων να έχει τιμές μικρότερες ή ίσες της τιμής της διάρκεσος. Στην περίπτωση του πληθυσμού η διάρκεσος συμβολίζεται με M και η αντίστοιχη δειγματική διάρκεσος συμβολίζεται με m . Η διάρκεσος εκφράζεται σε μονάδες μέτρησης της μεταβλητής. Η διάρκεσος είναι ένα μέτρο με ιδιαίτερη σημασία και χρησιμότητα στην ανάλυση ποσών φαινομένων οικονομικών, κοινωνικών, δημογραφικών κλπ. Για παράδειγμα, ο "διάρκεσος μισθός" μπορεί να θεωρηθεί ως ο "μέσος αμοιβάμενος εργαζόμενος" υπό την έννοια ότι οι μισθοί απασχολούμενων του κλάδου βρίσκονται σε κατώτερη και οι υπόλοιποι μισθοί σε ανώτερη μισθολογική θέση από αυτόν.

Για να βρούμε τη διάρκεσο διατάσσουμε τα αρχικά δεδομένα από την μικρότερη μέχρι την μεγαλύτερη τιμή. Εντοπίζουμε τη θέση της διάρκεσος ως εξής: Θέση διάρκεσος = $\frac{n+1}{2}$. Η τιμή της παρατήρησης του διατεταγμένου δείγματος που κατέχει αυτή τη θέση είναι η διάρκεσος (m) των δεδομένων μας.

Παραδείγματα

(α) Περισσότερο πλήθος παρατηρήσεων

Βάρος 7 ατόμων σε κιλά, $n=7$

80 65 60 70 67 78 68

Διατεταγμένο δείγμα: 60 65 67 68 70 78 80

$$\text{Θέση διάρκεσος} \quad \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$$

Η τέταρτη παρατήρηση του διατεταγμένου δείγματος έχει

τιμή 68. Άρα $m=68$ κιλά. Πριν από αυτήν την τιμή προηγούμενη -62-
νται 3 παρατηρήσεις και μετά από αυτήν έπονται 3 παρατη-
ρήσεις του διατεταγμένου δείγματος.

(β) Άρτιο πλήθος παρατηρήσεων

Βάρη 8 ατόμων σε κιλά

82 65 62 70 67 78 79 81

Διατεταγμένο δείγμα

62 65 67 70 78 79 81 82

Θέση διαμέσου $\frac{n+1}{2} = \frac{8+1}{2} = \frac{9}{2} = 4,5$, Η διάμεσος βρίσκεται

μεταξύ 4^{ης} και 5^{ης} παρατήρησης του διατεταγμένου δείγματος

(δηλ. η τιμή πρέπει να είναι μικρότερη των 78 κιλών και με-
γαλύτερη των 70 κιλών). Σ' αυτήν την περίπτωση θεω-
ρούμε συμβατικά ως τιμή της διαμέσου το ημίαθροισμα
των τιμών των δύο αυτών παρατηρήσεων δηλαδή:

$m = \frac{70+78}{2} = 74$ κιλά. Η προϋπάρχουσα τιμή της διαμέσου (=74)

δεν είναι παρατηρηθείσα τιμή είναι όμως δυνατή τιμή
της εξεταζόμενης μεταβλητής.

(γ) Έστω τα εξής δεδομένα ($n=9$)

6 6 2 5 6 7 1 4 8

Διατεταγμένο δείγμα: 1 2 4 5 6 6 6 7 8

Τα παραπάνω δεδομένα έχουν:

Επικρατούσα τιμή $\tau=6$, διάρκεια $m=6$, αριθμητικό μέσο

$\bar{x}=5$. Τροποποιούμε τα δεδομένα αντικαθιστώντας την τε-

λευταία παρατήρηση με την τιμή 857.

Νέα δεδομένα: 6 6 2 5 6 7 1 4 857

Διατεταγμένο νέο δείγμα 1 2 4 5 6 6 6 7 857

Τα παραπάνω στοιχεία έχουν την ίδια επικρατούσα τιμή $\tau=6$, την ίδια διάρκεια $m=6$ αλλά έχουν αριθμητικό μέσο $\bar{x}=100$, δηλαδή η τιμή της διάρκειας όπως και της επικρατούσας τιμής δεν επηρεάστηκε από την ύπαρξη της αυραίας τιμής ($=857$) σε αντίθεση με την τιμή του αριθμητικού μέσου η οποία αυξήθηκε κατά 95 μονάδες.

Στην περίπτωση οραδοποιημένων δεδομένων υπολογίζουμε τις αθροιστικές συχνότητες της κατανομής F_i . Εντοπίζουμε τη θέση της διάρκειας ως το ήμισυ των διαθέσιμων παρατηρήσεων δηλαδή $\frac{N}{2}$. Εντοπίζουμε την i-οστή τάξη της κατανομής στην οποία το ήμισυ των παρατηρήσεων δεν υπερβαίνει την αντίστοιχη συχνότητα (δηλαδή την τάξη της κατανομής για την οποία $\frac{N}{2} \leq F_i$). Εφαρμόζουμε τον τύπο:
$$M = L_i + \frac{\delta_i}{f_i} \left(\frac{N}{2} - F_{i-1} \right), \text{ όπου}$$

M = η τιμή της διάρκειας, i = η τάξη στα όρια της οποίας περιέχεται η διάρκεια, δ_i = το πλάτος της τάξης στην οποία εντοπίζεται η διάρκεια, f_i = η απλή συχνότητα της τάξης στην οποία εντοπίζεται η διάρκεια, L_i = το κατώτερο όριο τάξης στην οποία εντοπίζεται η διάρκεια, F_{i-1} = η αθροιστική συχνότητα της προηγούμενης τάξης από αυτήν στην οποία εντοπίζεται η διάρκεια.

Παράδειγμα δίνεται ο αριθμός των γεννήσεων για το 1988 που καταγράφηκαν στις αγροτικές περιοχές της Ελλάδας με σκοπό να υπολογιστεί η διάρκεια μητρία της μητέρας.

Έχουμε τον ακόλουθο πίνακα:

Τάξεις Ηλικιών	# γεννήσεων (f_i)	Αθροιστικές Συχνότητες (F_i)	-64-
μέχρι 15	12	12	
15-20	3.558	3570	
20-25	10.338	13908	
25-30	6.686	20594	
30-35	2904	23498	
35-40	1024	24522	
40-45	242	24764	
45-50	35	24799	
50 και άνω	3	24802	
Σύνολο	24802		

Θέση διαμέσου $\frac{N}{2} = \frac{24802}{2} = 12401$. Η διάμεσος εντοπίζεται

στην τρίτη τάξη της κατανομής διότι σ' αυτήν είναι:

$\frac{N}{2} = 12401 < F_3 = 13908$. Η τιμή της διαμέσου θα πρέπει

να είναι $20 < M < 25$ και υπολογίζεται ως εξής:

$$M = L_3 + \frac{\delta_3}{f_3} \left(\frac{N}{2} - F_{3-1} \right) = 20 + \frac{5}{10338} (12401 - 3570)$$

$= 24,27$ έτη, δηλαδή το έτος 1988 οι μισές γεννήσεις πρα-

γματοποιήθηκαν από μητέρες ηλικίας μέχρι 24,27 ετών

και οι υπόλοιπες μισές από μητέρες οι οποίες ήταν ηλι-

κίας 24,27 ετών και άνω.

§5.3 Τεταρτηρόρια (Quartiles) Αυτά υποδιαιρούν το σύνολο των διαθέσιμων μετρήσεων σε 4 ισοπαμθές ορά-
δες κάθε μία από τις οποίες περιλαμβάνει το 25% των διατεταγμένων παρατηρήσεων.

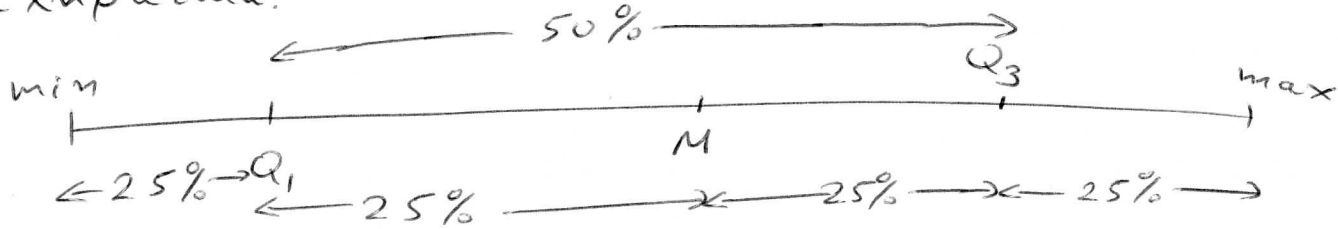
Το πρώτο τεταρτηρόριο (first quartile, Q_1) είναι η τιμή της μεταβλητής μέχρι την οποία περιλαμβάνεται το 25%

(το $\frac{1}{4}$) των διατεταγμένων τιμών του δείγματος.

Το τρίτο τεταρτηγόριο (third quartile, Q_3) είναι η τιμή της μεταβλητής μέχρι την οποία περιλαμβάνεται το 75% (ή $\frac{3}{4}$) των διατεταγμένων τιμών του δείγματος.

Το δεύτερο τεταρτηγόριο (second quartile, Q_2) συμπίπτει με τη διάμεσο της κατανομής.

Σχηματικά:



Άρα, από την τιμή του Q_1 και άνω περιλαμβάνεται το 75% των διατεταγμένων δείγματος. Από την τιμή του Q_3 και άνω περιλαμβάνεται το 25% του διατεταγμένου δείγματος και μεταξύ των τιμών Q_1 και Q_3 περιλαμβάνεται το 50% των διατεταγμένων τιμών του δείγματος. Η διαφορά ανάμεσα στην τιμή του τρίτου και του πρώτου τεταρτηγρίου καλείται ενδο-τεταρτηγοριακό εύρος (interquartile range) και συμβολίζεται με IQR δηλαδή:

$$IQR = Q_3 - Q_1$$

Εκφράζεται σε μονάδες μέτρησης της μεταβλητής και δίνει το κεντρικό διάστημα μέσα στο οποίο ερπίζει το 50% των διατεταγμένων τιμών του δείγματος. Το ήμισυ του καλείται ημι-ενδοτεταρτηγοριακό εύρος (semi-interquartile range) και συμβολίζεται με

$$Q: \quad Q = \frac{1}{2} IQR = \frac{1}{2} (Q_3 - Q_1).$$

Η τιμή του τεταρτηφορίου εύρους εξαρτάται από το πόσο διεσπαρμένες είναι οι τιμές των παρατηρήσεων. Ανταναγκά τη διακύμανση μόνο των κεντρικών τιμών του δείγματος σε αντίθεση με τη διασπορά (ή τη διακύμανση) που θα δοίμε στο επόμενο κεφάλαιο και η οποία ανταναγκά τη υόρανση όλων των τιμών του δείγματος. Το IQR χρησιμοποιεί κυρίως για τη σύγκριση της διακύμανσης των τιμών μεταξύ δύο κατανομών.