2.3.1 Παράμετροι θέσης

Με τις παραμέτρους θέσης προσπαθούμε να καθορίσουμε τη κεντρική θέση της κατανομής, δηλαδή στο σημείο που αντιστοιχεί στην τιμή της μεταβλητής γύρω από την οποία τείνουν να συγκεντρωθούν οι μεταβλητές του πληθυσμού.

2.3.1.1 Αριθμητικός μέσος

Ο αριθμητικός μέσος μίας μεταβλητής (συμβολισμός $\overline{X}$) ορίζεται ως το πηλίκο του αθροίσματος των τιμών της δια το πλήθος τους, δηλαδή για τη μεταβλητή X με τιμές $X_1, X_2, \ldots, X_n$ έχουμε:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Ο αριθμητικός μέσος των τιμών ενός δείγματος χρησιμοποιείται για την εκτίμηση της μέσης τιμής των τιμών του πληθυσμού (συμβολισμός μ).

Παράδειγμα 2.1:

Από ένα δείγμα 16 ατόμων καταγράψαμε τις παρακάτω ηλικίες: 33, 36, 5, 8, 36, 30, 12, 13, 4, 34, 40, 65, 4, 2,5, 2,5, και 43. Να βρεθεί η μέση τιμή της ηλικίας στο δείγμα.

$$\overline{X} = \frac{\sum_{i=1}^{16} X_i}{16} = \frac{368}{16} = 23$$

Στην περίπτωση που θέλουμε να υπολογίσουμε τον αριθμητικό μέσο μίας διακριτής μεταβλητής που λαμβάνει τις τιμές X_1, $X_2, \ldots, X_k$ με αντίστοιχες συχνότητες n_1, $n_2, \ldots,$ n_k, μπορούμε να χρησιμοποιήσουμε και τον τύπο:

$$\overline{X} = \frac{\sum_{i=1}^{k} n_i X_i}{\sum_{i=1}^{k} n_i} = \frac{n_1 X_1 + n_2 X_2 + \cdots + n_k X_k}{n_1 + n_2 + \cdots + n_k}$$

Στη περίπτωση που έχουμε μία συνεχή μεταβλητή οι τιμές της οποίας έχουν ταξινομηθεί σε κλάσεις θα θεωρήσουμε ότι οι παρατηρήσεις κάθε ταξικού διαστήματος είναι συγκεντρωμένες στο μέσο του. Έτσι, αν η συχνότητα κάθε της κλάσης $[e_{i-1}, e_i)$ είναι f_i έχουμε:

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} = \frac{f_1 X_1 + f_2 X_2 + \cdots + f_k X_k}{f_1 + f_2 + \cdots + f_k}, \quad X_i = \frac{e_{i-1} + e_i}{2}$$

Παράδειγμα 2.2

Να βρεθεί η μέση τιμή της ηλικίας του δείγματος που παρουσιάζεται στον πίνακα 2.8.

Στον πίνακα 2.8 θα προσθέσουμε δύο νέες στήλες: Την στήλη με επικεφαλίδα X_i στην οποία θα καταγράψουμε το κέντρο της κάθε κλάσης και τη στήλη f_iX_i στην οποία θα υπολογίσουμε το γινόμενο της συχνότητας επί του κέντρου της κάθε κλάσης.

Πίνακας 2.13 Μέσης τιμή ομαδοποιημένης ποσοτικής μεταβλητής

ΗΛΙΚΙΑ	f_i	X_i	f_iX_i
[0, 6)	166	3	498
[6, 12)	146	9	1314
[12 ,18)	178	15	2670
[18, 24)	229	21	4809
[24, 30)	229	27	6183
[30, 36)	221	33	7293
[36, 42)	217	39	8463
[42, 48)	213	45	9585
[48, 54)	213	51	10863
[54, 60)	139	57	7923
[60, 66)	115	63	7245
[66, 72)	66	69	4554
[72, 78)	52	75	3900
[78, 84)	23	81	1863
[84, 90)	3	87	261
[90, 96)	4	93	372
Σύνολο	2214		77796

Από τον παραπάνω πίνακα εύκολα καταλήγουμε στη μέση τιμή της ηλικίας:

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} = \frac{77796}{2214} = 35.14$$

2.3.1.2 Ιδιότητες αριθμητικού μέσου

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right) = 0$$

Αν $X_1 = X_2 = \cdots = X_n = \alpha$ τότε $\overline{X} = \alpha$

Αν $\alpha \le X_i \le \beta$ τότε $\alpha \le \overline{X} \le \beta$

Αν $Y = \alpha + \beta X$ τότε $\overline{Y} = \alpha + \beta \overline{X}$

$$\sum_{i=1}^{n}\left(X_i - \alpha\right)^2 \ge \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

2.3.1.3 Σταθμικός αριθμητικός μέσος

Αν στις τιμές $X_1, X_2, \ldots, X_k$ της μεταβλητής X αντιστοιχούν οι συντελεστές βαρύτητας $\Pi_1, \Pi_2, \ldots, \Pi_k$ ο σταθμικός αριθμητικός μέσος υπολογίζεται με τον τύπο:

$$\overline{X} = \frac{\sum_{i=1}^{k} \Pi_i X_i}{\sum_{i=1}^{k} \Pi_i} = \frac{\Pi_1 X_1 + \Pi_2 X_2 + \cdots + \Pi_k X_k}{\Pi_1 + \Pi_2 + \cdots + \Pi_k}$$

Εξαιρετικά μεγάλες ή εξαιρετικά μικρές τιμές στα δεδομένα επηρεάζουν τον αριθμητικό μέσο έτσι ώστε να μη σηματοδοτηθεί ικανοποιητικά τη θέση του κυρίως όγκου των δεδομένων. Γι' αυτό σε πολλές πρακτικές εφαρμογές υπολογίζεται ο αποκομμένος μέσος, ο οποίος υπολογίζεται αφαιρώντας από τα δεδομένα ορισμένο ποσοστό μεγαλύτερων και μικρότερων τιμών. Έτσι π.χ. ο κατά 20% αποκομμένος μέσος είναι ο αριθμητικός μέσος ο οποίος υπολογίζεται αν αγνοήσουμε το 20% των μεγαλύτερων και το 20% των μικρότερων τιμών. Ομοίως, ο κατά 10% αποκομμένος μέσος είναι ο μέσος που υπολογίζεται μετά την αφαίρεση του 10% των μεγαλύτερων και του 10% των μικρότερων τιμών των δεδομένων.

2.3.1.4 Διάμεση τιμή

Η διάμεση τιμή M των τιμών $X_1, X_2, \ldots, X_n$ μίας μεταβλητής X είναι εκείνη η τιμή της μεταβλητής που πριν και μετά της υπάρχει ο ίδιος αριθμός τιμών όταν αυτές έχουν διαταχθεί σε αύξουσα ή φθίνουσα σειρά. Δηλαδή το 50% των τιμών της μεταβλητής είναι μικρότερο και το 50% μεγαλύτερο της διάμεσης τιμής M.

Αν το πλήθος n των τιμών της μεταβλητής είναι περιττός αριθμός τότε η διάμεσος βρίσκεται στη θέση $(n+1)/2$, ενώ αν είναι άρτιος τότε η διάμεσος βρίσκεται στη θέση $n/2+1$.

Παράδειγμα 2.3

Να βρεθεί η διάμεσος των δεδομένων του παραδείγματος 2.1

Αρχικά ταξινομούμε τα δεδομένα σε αύξουσα σειρά: 2.5, 2.5, 4, 4, 5, 8, 12, 13, 30, 33, 34, 36, 36, 40, 43, 65. Επειδή το πλήθος των δεδομένων είναι άρτιος αριθμός (n=16) θα επιλέξουμε ως διάμεσο το ένατο στοιχείο, άρα M=30.

Εάν έχουμε υπολογίσει την απόλυτη ή τη σχετική αθροιστική συχνότητα τότε μπορούμε να υπολογίσουμε τη διάμεσο ως τη τετμημένη των τιμών $F(n/2)$ και $F/n(0,5)$ αντίστοιχα. Αν θέλουμε να υπολογίσουμε τη διάμεσο M μίας διακριτής μεταβλητής X τότε επιλέγουμε εκείνη η τιμή που αντιστοιχεί στη μεγαλύτερη από τις δύο αθροιστικές συχνότητες F_i (F_i/n) μεταξύ των οποίων περιέχεται ο αριθμός $n/2$ (0,5).

$$M=X_i \text{ για } F_{i-1}< n/2 <F_i$$
$$M=X_i \text{ για } F_{i-1}/n< 0,5 <F_i/n$$

Παράδειγμα 2.4

Να βρεθεί η διάμεσος των δεδομένων του πίνακα 2.11

Μπορούμε να δουλέψουμε είτε με τις αθροιστικές συχνότητες, οπότε θα αναζητήσουμε τις κλάσεις των οποίων οι αθροιστικές συχνότητες περιβάλουν τον αριθμό 78/2=39 είτε με τις αθροιστικές σχετικές συχνότητες, οπότε θα αναζητήσουμε τις κλάσεις των οποίων οι αθροιστικές συχνότητες περιβάλουν τον αριθμό 0.5. Προφανώς και στις δύο περιπτώσεις θα καταλήξουμε στις ίδιες κλάσεις.

Η (σχετική) αθροιστική συχνότητα (0.5) 39 βρίσκεται ανάμεσα στις (σχετικές) αθροιστικές συχνότητες (0.47) 37 και (0.67) 52 οι οποίες αντιστοιχούν στις κλάσεις 5 και 6. Συνεπώς ως διάμεσο θα επιλέξουμε τη μεγαλύτερη τιμή, δηλαδή M=6.

Πίνακας 2.14 Υπολογισμός διαμέσου σε διακριτή μεταβλητή

ΜΗΝΕΣ ΑΝΕΡΓΙΑΣ	f_i	f_i/n	F_i	F_i/n
1	7	0.09	7	0.09
2	8	0.10	15	0.19
3	11	0.14	26	0.33
4	7	0.09	33	0.42
5	**4**	**0.05**	**37**	**0.47**
6	**15**	**0.19**	**52**	**0.67**
7	0	0.00	52	0.67
8	8	0.10	60	0.77
9	1	0.01	61	0.78
10	9	0.12	70	0.90
11	2	0.03	72	0.92
12	6	0.08	78	1.00
Σύνολο	78			

Στη περίπτωση που έχουμε μία συνεχή μεταβλητή οι τιμές της οποίας έχουν ταξινομηθεί σε κλάσεις, θα θεωρήσουμε ότι οι παρατηρήσεις κάθε ταξικού διαστήματος είναι ομοιόμορφα κατανεμημένες σε αυτό. Έτσι, αν η συχνότητα της κάθε κλάσης $[e_{i-1}, e_i)$ είναι f_i έχουμε:

$$M = e_{i-1} + \frac{e_i - e_{i-1}}{f_i}\left(\frac{n}{2} - F_{i-1}\right)$$

Παράδειγμα 2.5

Να βρεθεί η μέση τιμή της ηλικίας του δείγματος που παρουσιάζεται στον πίνακα 2.8.

Πίνακας 2.15 Διάμεσος ομαδοποιημένης ποσοτικής μεταβλητής

ΗΛΙΚΙΑ	f_i	F_i
[0, 6)	166	166
[6, 12)	146	312
[12 ,18)	178	490
[18, 24)	229	719
[24, 30)	**229**	**948**
[30, 36)	**221**	**1169**
[36, 42)	217	1386
[42, 48)	213	1599
[48, 54)	213	1812
[54, 60)	139	1951
[60, 66)	115	2066
[66, 72)	66	2132
[72, 78)	52	2184
[78, 84)	23	2207
[84, 90)	3	2210
[90, 96)	4	2214
Σύνολο	2214	

Η αθροιστική συχνότητα 2214/2=1107 βρίσκεται ανάμεσα στις αθροιστικές συχνότητες 948 και 1169 οι οποίες αντιστοιχούν στις κλάσεις [24, 30) και [30, 36). Συνεπώς η διάμεσος υπολογίζεται ως εξής:

$$M = e_{i-1} + \frac{e_i - e_{i-1}}{f_i}\left(\frac{n}{2} - F_{i-1}\right) = 30 + \frac{36-30}{221}(1107-948) = 34.32$$

2.3.1.5 Ιδιότητες διάμεσης τιμής

Η διάμεση τιμή, σε αντίθεση με τον αριθμητικό μέσο, δεν επηρεάζεται από τις ακραίες τιμές των παρατηρήσεων. Αυτό οφείλεται στο ότι για τον υπολογισμό της λαμβάνεται υπόψη κυρίως η διάταξη των δεδομένων και όχι το πραγματικό τους μέγεθος.

2.3.1.6 Τεταρτημόρια και εκατοστιαία σημεία

Με την ίδια λογική που χρησιμοποιήθηκε με τη διάμεσο για το χωρισμό του πληθυσμού σε δύο ίσους πληθυσμούς, είναι δυνατό να προσδιοριστούν οι παρακάτω τιμές μιας μεταβλητής:

Δεκατημόρια: $D_1, D_2, \ldots, D_9$ που χωρίζουν το πληθυσμό σε 10 ίσους πληθυσμούς.

Τεταρτημόρια: Q_1, Q_2, Q_3 που χωρίζουν το πληθυσμό σε 4 ίσους πληθυσμούς.

Γενικότερα όταν οι παρατηρήσεις διαταχτούν κατά αύξουσα σειρά, τότε για οποιοδήποτε ποσοστό p μπορούμε να προσδιορίσουμε την τιμή για την οποία ισχύει: ποσοστό το πολύ ίσο με p % των παρατηρήσεων είναι μικρότερο της και, αντίστοιχα, ποσοστό το πολύ ίσο με (1-p)% είναι μεγαλύτερο της. Η παρατήρηση αυτή ονομάζεται p -εκατοστιαίο σημείο και βρίσκεται στη θέση $(n+1)p/100$ όπου n το πλήθος των παρατηρήσεων.

Είναι προφανές ότι για να προσδιορίσουμε τη θέση αυτή είναι απαραίτητο οι παρατηρήσεις να έχουν διαταχθεί σε αύξουσα σειρά. Σύμφωνα και με τον ορισμό της διαμέσου, το 50-οστό εκατοστιαίο σημείο είναι η διάμεσος, δηλαδή $M=Q_2$=50-*percentile*. Πράγματι, όταν το πλήθος των παρατηρήσεων είναι άρτιο, τότε το 50% των παρατηρήσεων είναι μικρότερες απ' αυτήν και, αντίστοιχα, το 50% είναι μεγαλύτερες απ' αυτήν. Όταν το πλήθος των παρατηρήσεων είναι περιττό, το ποσοστό των παρατηρήσεων που είναι μικρότερες (μεγαλύτερες) από την διάμεσο είναι λίγο μεγαλύτερο (μικρότερο) από το 50%.

Στη περίπτωση που έχουμε μία συνεχή μεταβλητή οι τιμές της οποίας έχουν ταξινομηθεί σε κλάσεις, θα θεωρήσουμε ότι οι παρατηρήσεις κάθε ταξικού διαστήματος είναι ομοιόμορφα κατανεμημένες σε αυτό. Έτσι, αν η συχνότητα της κάθε κλάσης $[e_{i-1}, e_i)$ είναι f_i έχουμε για το p -εκατοστιαίο σημείο:

$$p - percentile = e_{i-1} + \frac{e_i - e_{i-1}}{f_i}\left(p\frac{n}{100} - F_{i-1}\right)$$

Για παράδειγμα το πρώτο και το τρίτο τεταρτημόριο για τα οποία ισχύει Q_1=25-*percentile* και Q_3=75-*percentile* υπολογίζονται ως εξής:

$$Q_1 = e_{i-1} + \frac{e_i - e_{i-1}}{f_i}\left(\frac{n}{4} - F_{i-1}\right), \quad Q_3 = e_{i-1} + \frac{e_i - e_{i-1}}{f_i}\left(3\frac{n}{4} - F_{i-1}\right)$$

2.3.1.7 Επικρατέστερη τιμή

Η επικρατέστερη τιμή των τιμών μίας μεταβλητής είναι εκείνη η τιμή στην οποία αντιστοιχεί η μεγαλύτερη συχνότητα παρατηρήσεων.

2.3.1.8 Σχέση μέσης, διάμεσης και επικρατέστερης τιμής

Όταν τα δεδομένα είναι ονομαστικά τότε η επικρατούσα τιμή δηλαδή αυτή με τη μεγαλύτερη συχνότητα είναι το μόνο μέτρο θέσης που μπορούμε να χρησιμοποιήσουμε. Όταν τα δεδομένα είναι διατακτικά έχει νόημα να προσδιορίσουμε εκτός από την επικρατούσα και την διάμεσο τιμή. Όταν τα δεδομένα είναι σε διαστημική ή αναλογική κλίμακα μπορούμε να χρησιμοποιήσουμε την επικρατούσα τιμή, τη διάμεσο και τον αριθμητικό μέσο. Το μέτρο που τελικά θα επιλέξουμε εξαρτάται από τη φύση του προβλήματος και τους σκοπούς της έρευνας. Έτσι π.χ. αν στα δεδομένα έχουμε λίγες, διακριτές τιμές, οι οποίες επαναλαμβάνονται με κάποιες συχνότητες τότε η επικρατούσα τιμή και η διάμεσος δίνουν καλύτερη πληροφόρηση στον μέσο αναγνώστη από τον αριθμητικό μέσο, ο οποίος μπορεί να πάρει και μια τιμή η οποία είναι μη παρατηρήσιμη. Έτσι π.χ. αν τα δεδομένα είναι παρατηρήσεις του αριθμού παιδιών ανά οικογένεια ο μέσος αναγνώστης αντιλαμβάνεται καλύτερα τις προτάσεις «η τυπική οικογένεια έχει 2 παιδιά» και «οι μισές οικογένειες έχουν μέχρι ένα παιδί» από την πρόταση «ο μέσος αριθμός ανά οικογένεια ισούται με 1.55». Ο αριθμητικός μέσος όμως μπορεί να χρησιμοποιηθεί σε περαιτέρω στατιστική ανάλυση.

Γενικότερα, από άποψη μαθηματικής δομής τα ονομαστικά δεδομένα βρίσκονται στο κατώτερο επίπεδο μέτρησης, ακολουθούν τα διατακτικά και τέλος τα ποσοτικά. Οι στατιστικές μέθοδοι που έχουν αναπτυχθεί για τα δεδομένα ορισμένου τύπου μπορούν να χρησιμοποιηθούν και στα δεδομένα ανώτερης κλίμακας αλλά όχι στις κατώτερες.

Τέλος σημειώνουμε ότι στις συμμετρικές κατανομές οι τιμές των τριών παραμέτρων συμπίπτουν. Στις ασύμμετρες κατανομές οι τιμές των παραμέτρων θέσης διαφέρουν και η διάμεσος βρίσκεται μεταξύ των άλλων δύο. Στην περίπτωση θετικής ασυμμετρίας, ο αριθμητικός μέσος είναι η μεγαλύτερη των τριών παραμέτρων θέσης. Στην περίπτωση αρνητική ασυμμετρίας η επικρατέστερη τιμή είναι η μεγαλύτερη των τριών παραμέτρων θέσης.