

Κεφάλαιο 5

Οι δείκτες διασποράς

1

Ένα παράδειγμα εργασίας

Ένας καθηγητής μαθηματικών έδωσε σε δύο τμήματα μιας τάξης του σχολείου του το ίδιο τεστ. Η επίδοση των μαθητών του κάθε τμήματος (όπως μετρήθηκε με τη χρήση μιας εικοσαβάθμιας κλίμακας) παρουσιάζεται στον Πίνακα:

Τμήμα Α				Τμήμα Β			
17	18	19	14	20	16	14	17
18	17	13	16	17	20	9	16
20	13	17	17	19	16	17	20
14	17	15	18	15	17	9	16
16	12	19	17	17	10	20	17
14				19			

Ποιο από τα δύο τμήματα είχε καλύτερη επίδοση;

2

- ▣ Οι επιδόσεις των μαθητών των δύο τμημάτων δεν είναι οι ίδιες...
- ▣ Ωστόσο, οι τιμές και των τριών δεικτών κεντρικής τάσης είναι ακριβώς οι ίδιες μεταξύ των δύο ομάδων!
- ▣ Η **δεσπόζουσα τιμή** και για τα δύο τμήματα είναι η τιμή **17** (η συχνότητά της είναι 6 και στα δύο τμήματα)
- ▣ Η **διάμεσος** και στις δύο ομάδες είναι **16,75**
- ▣ Τέλος, ο **αριθμητικός μέσος όρος** και των δύο ομάδων είναι επίσης ο ίδιος: **16,24**
- ▣ Άρα;...

3

Διάγραμμα μίσχου και φύλλων

Τμήμα Α	2	Τμήμα Β
0	2	0000
99	1	99
888	1	
777777	1	777777
66	1	6666
5	1	55
444	1	4
33	1	
2	1	
	1	
	1	0
	0	99

4

Διασπορά ή σκεδασμός

Γενικός στατιστικός όρος για κάθε μέτρηση που αναφέρεται στη διακύμανση των τιμών μιας ομάδας τιμών:

- ▣ **εύρος** (range),
- ▣ **ενδοτεταρτημοριακό εύρος** (interquartile range),
- ▣ **μέση αριθμητική απόκλιση** (mean deviation), και
- ▣ **τυπική απόκλιση** (standard deviation).

5

Εύρος

Η διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής της κατανομής

Πλεονεκτήματα:

- Είναι πολύ εύκολο στον υπολογισμό του
- Περιλαμβάνει και τις ακραίες τιμές της κατανομής

Μειονεκτήματα:

- ▣ Αλλοιώνεται από τις ακραίες τιμές με αποτέλεσμα, σε πολλές περιπτώσεις, να μην παρουσιάζει μια αντιπροσωπευτική εικόνα της διασποράς της κατανομής
- ▣ Δεν παρέχει καμιά πληροφορία σχετικά με τη διασπορά των τιμών μεταξύ των άκρων της κατανομής. Για παράδειγμα, δεν μας λέει τίποτα για τη διασπορά των τιμών της κατανομής γύρω από το μέσο όρο

6

Ενδοτεταρτημοριακό εύρος

Ενδοτεταρτημοριακό εύρος: Το εύρος του κεντρικού 50% των τιμών μιας κατανομής

Τεταρτημόρια: Τα σημεία που χωρίζουν την κατανομή σε τέσσερα ίσα μέρη:

- 1ο τεταρτημόριο ή 25ο
- 2ο τεταρτημόριο ή 50ο ή διάμεσος
- 3ο τεταρτημόριο ή 75ο

Ο τύπος υπολογισμού των τεταρτημορίων:

$$\text{Τιμή τεταρτημορίου} = L + \frac{P - F}{f_m}$$

Όπου,

L = το ακριβές ανώτερο πραγματικό όριο της προηγούμενης τιμής,

P = το ποσοστό των τιμών της κατανομής που βρίσκονται κάτω από το συγκεκριμένο τεταρτημόριο,

F = η αθροιστική σχετική συχνότητα της προηγούμενης τιμής, και

f_m = η επιμέρους σχετική συχνότητα της επόμενης τιμής.

7

Ενδοτεταρτημοριακό εύρος

■ Το ενδοτεταρτημοριακό εύρος ισούται με τη διαφορά μεταξύ του 1ου και του 3ου τεταρτημορίου.

■ Εφαρμόζουμε τον τύπο που είδαμε στην προηγούμενη διαφάνεια και υπολογίζουμε την τιμή των δύο τεταρτημορίων:

■ 1ο τεταρτημόριο = 14,26

■ 3ο τεταρτημόριο = 17,75

Επομένως:

$$1\text{ο τεταρτημόριο} - 3\text{ο τεταρτημόριο} = 14,26 - 17,75 = -3,49$$

8

Ενδοτεταρτημοριακό εύρος

Πλεονεκτήματα:

- Δεν επηρεάζεται από τις ακραίες τιμές
- Είναι σχετικά εύκολο στον υπολογισμό του
- Είναι αντιπροσωπευτικό των κεντρικών τιμών της κατανομής

Μειονεκτήματα:

- Δεν λαμβάνει υπόψη τις ακραίες τιμές της κατανομής
- Όπως και το εύρος, δεν επιτρέπει την ακριβή ερμηνεία μιας συγκεκριμένης τιμής της κατανομής
- Δεν είναι ακριβές όταν τα δεδομένα είναι ομαδοποιημένα κατά μεγάλα διαστήματα τιμών
- Όπως και η διάμεσος, δεν περιγράφει καμιά από τις παραμέτρους, οι οποίες είναι βασικές για την επαγωγική στατιστική

9

Μέση αριθμητική απόκλιση

Απόκλιση: Η αριθμητική διαφορά ανάμεσα σε μια τιμή και στο μέσο όρο της κατανομής στην οποία ανήκει η συγκεκριμένη τιμή

Ο τύπος υπολογισμού της απόκλισης:

$$d_i = X_i - \bar{X}$$

10

Μέση αριθμητική απόκλιση

Τιμές	$X_i - \bar{X}$	d	Απόλυτη επιμέρους συχνότητα (f)	fd
20	20-16.24	3.76	1	3.76
19	19-16.24	2.76	2	5.52
18	18-16.24	1.76	3	5.28
17	17-16.24	0.76	6	4.56
16	16-16.24	-0.24	2	-0.48
15	15-16.24	-1.24	1	-1.24
14	14-16.24	-2.24	3	-6.72
13	13-16.24	-3.24	2	-6.48
12	12-16.24	-4.24	1	-4.24
Σύνολο			21	0

11

Μέση αριθμητική απόκλιση

Ο τύπος υπολογισμού της μέσης αριθμητικής απόκλισης:

$$MD = \frac{\sum |X - \bar{X}|}{N}$$

12

Τυπική απόκλιση

Διακύμανση: Ο μέσος όρος των τετραγώνων των αποκλίσεων όλων των τιμών μιας κατανομής

Τυπική απόκλιση: Δείκτης διασποράς αντιπροσωπευτικός των αποκλίσεων μιας ομάδας τιμών από το μέσο όρο

Ο τύπος υπολογισμού της τυπικής απόκλισης:

$$s = \sqrt{\frac{\sum d^2}{N}}$$

13

Τυπική απόκλιση

Πλεονεκτήματα:

- Μπορεί να χρησιμοποιηθεί για τον υπολογισμό των παραμέτρων του πληθυσμού
- Λαμβάνει υπόψη όλες τις τιμές της κατανομής
- Είναι ο πιο ευαίσθητος από τους δείκτες διασποράς

Μειονεκτήματα:

- Ο υπολογισμός της είναι σχετικά πιο περίπλοκος σε σχέση με τους υπόλοιπους δείκτες διασποράς
- Είναι πολύ ευαίσθητη στις ακραίες τιμές της κατανομής

14

Επιλέγοντας τον κατάλληλο δείκτη διασποράς

- Μεγάλη σημασία έχουν οι ακραίες τιμές στην επιλογή του κατάλληλου δείκτη.
- Ωστόσο, πρέπει να εξετάσουμε και το επίπεδο μέτρησης που έχουμε επιτύχει.
- Έτσι, στην ιεραρχική κλίμακα μέτρησης, μεγαλύτεροι αριθμοί δείχνουν μεγαλύτερη ποσότητα από οτιδήποτε μετρείται, αλλά μεγαλύτερες και μικρότερες διαφορές μεταξύ των αριθμών μπορεί να μη δείχνουν μεγαλύτερες και μικρότερες διαφορές ανάμεσα στα πράγματα που μετρώνται. Σε μια τέτοια περίπτωση αρκεί ο υπολογισμός του εύρους.
- Σε κλίμακα ίσων διαστημάτων ή αναλογική κλίμακα, μεγάλες διαφορές στις μετρήσεις αντιστοιχούν πράγματι σε μεγάλες διαφορές στα πράγματα που μετρώνται. Σε αυτή την περίπτωση, και εφόσον δεν αναμένεται να υπάρχουν ακραίες τιμές, επιλέγουμε το μέσο όρο και την τυπική απόκλιση.

15

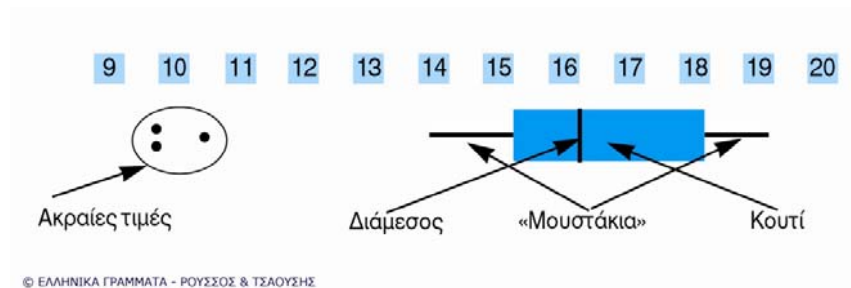
Επιλέγοντας τον κατάλληλο δείκτη διασποράς

- Αν είναι πιθανό να υπάρχουν ακραίες τιμές ή αν η μέτρηση είναι σε ιεραρχική κλίμακα, τότε πρέπει να χρησιμοποιούνται η διάμεσος και το ενδοτεταρτημοριακό εύρος. Η δεσπόζουσα τιμή και το εύρος μπορούν να χρησιμοποιηθούν αν επαρκεί μία κατά προσέγγιση εικόνα των τιμών του δείγματος.
- Ο συνηθέστερος τρόπος περιγραφής της διασποράς των τιμών μιας μεταβλητής είναι μέσω της τυπικής απόκλισης. Ο σημαντικότερος λόγος για τον οποίο προτιμάται η τυπική απόκλιση από τους υπόλοιπους δείκτες διασποράς είναι η δυνατότητα που προσφέρει να υπολογίσουμε παραμέτρους του πληθυσμού.

16

Διάγραμμα κουτιού

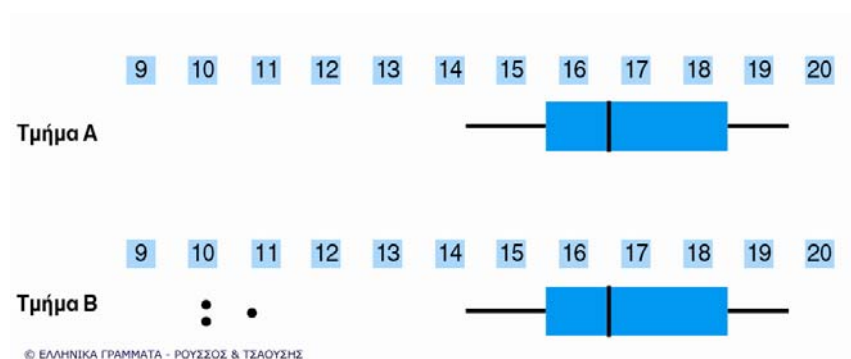
Διάγραμμα που χρησιμοποιείται για τη γραφική παράσταση της διασποράς ενός δείγματος



17

Διαγράμματα κουτιού του παραδείγματος

Διαγράμματα κουτιού για τις τιμές των δύο τμημάτων του παραδείγματος μας



18