



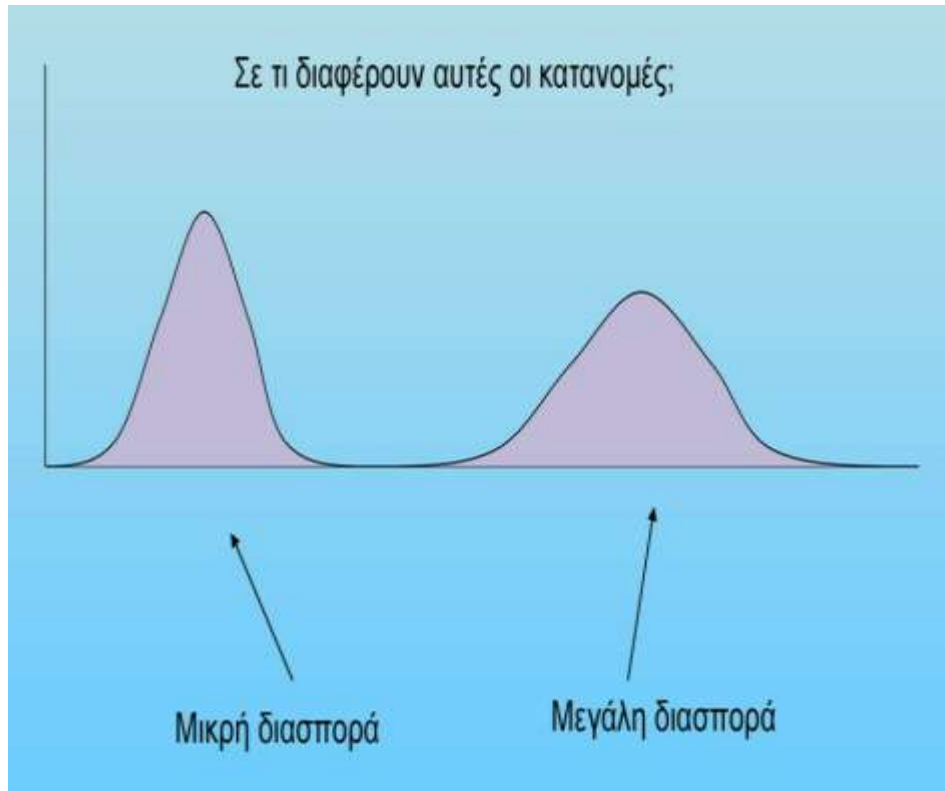
ΕΡΓΑΣΤΗΡΙΑ ΜΑΘΗΜΑΤΟΣ «Μέθοδοι Επικοινωνιακής Έρευνας II»

Εργαστήριο 3
Δείκτες διασποράς

ΠΕΡΙΓΡΑΜΜΑ ΠΑΡΟΥΣΙΑΣΗΣ

1. Δείκτες διασποράς.
 - A) Εύρος
 - B) Τυπική απόκλιση
2. Τυπική απόκλιση πληθυσμού.
3. Ο κανόνας 68-95-99.

1. ΔΕΙΚΤΕΣ ΔΙΑΣΠΟΡΑΣ



Αριθμητικές τιμές που δείχνουν το βαθμό διασποράς των τιμών μιας κατανομής γύρω από το «κέντρο» της.

Οι δείκτες κεντρικής τάσης (μέσος όρος, διάμεσος, δεσπόζουσα) δεν είναι αρκετοί από μόνοι τους για να περιγράψουν μια κατανομή.

Ένα παράδειγμα εργασίας

Ένας καθηγητής μαθηματικών έδωσε σε δύο τμήματα ενός φροντιστηρίου το ίδιο τεστ. Η επίδοση των μαθητών του κάθε τμήματος (όπως μετρήθηκε με τη χρήση 20βάθμιας κλίμακας) παρουσιάζεται στον Πίνακα:

Τμήμα Α		Τμήμα Β	
15	13	15	10
16	15	11	20
17	14	19	15

Ποιο από τα δύο τμήματα είχε καλύτερη επίδοση;

- Οι επιδόσεις των μαθητών των δύο τμημάτων **δεν είναι** ίδιες.
- Ωστόσο, οι τιμές των τριών δεικτών κεντρικής τάσης είναι ακριβώς οι ίδιες και στις δύο ομάδες!
- Η **δεσπόζουσα τιμή** και για τα δύο τμήματα είναι η τιμή **15**.
- Η **διάμεσος** και στις δύο ομάδες είναι ίση με **15**.
- Ο **μέσος όρος** και των δύο τμημάτων είναι επίσης ο ίδιος: **15**.
- Άρα;...

- Παρατηρούμε ότι το τμήμα Β έχει **περισσότερο διεσπαρμένες** τιμές.
- Παρατηρούνται, δηλαδή, μεγαλύτερες αποκλίσεις από τον μέσο όρο σε σχέση με το τμήμα Α (κάτι πολύ ενδιαφέρον για τον καθηγητή).
- Πρόκειται για μια διαφορά που κανένα από τα μέτρα κεντρικής τάσης δεν μπορεί να περιγράψει.

Ποιο από τα δύο τμήματα είναι δυσκολότερο να διδάξει ο καθηγητής και γιατί;

A) Εύρος (L)

- Ισούται με τη διαφορά της ελάχιστης τιμής από τη μέγιστη τιμή μιας κατανομής ($\max - \min$).
- **Προσοχή:** το εύρος είναι **μια συγκεκριμένη τιμή** και όχι το διάστημα μεταξύ δύο τιμών!

Στο προηγούμενο παράδειγμα ποιο είναι το εύρος των τιμών του τμήματος A και του τμήματος B;

Να βρεθεί το εύρος των παρατηρήσεων:

4, 3, 3, 2, 5, 8, 6, 6, 7, 10, 9, 12.

$$L = \max - \min = 12 - 2 = 10$$

Ποιο είναι το βασικό μειονέκτημα του εύρους (L) ως δείκτη διασποράς;

Π.χ. δύο κατανομές με την ίδια μέγιστη και ελάχιστη τιμή (άρα το ίδιο εύρος) μπορεί να έχουν πολύ διαφορετική διασπορά τιμών μεταξύ των δύο αυτών άκρων.

Κατανομή 1: **10** 13 14 15 16 17 18 **20 (L=10)**

Κατανομή 2: **10** 11 12 13 17 18 19 **20 (L=10)**

B) Τυπική απόκλιση (T.A. ή s)

- Είναι δείκτης διασποράς που λαμβάνει υπόψη του τις αποκλίσεις (αποστάσεις) **όλων των τιμών** μιας κατανομής από τον μέσο όρο.
- Είναι **ο πιο αντιπροσωπευτικός** δείκτης διασποράς.
- Είναι **κομβικής σημασίας** για την εφαρμογή τεχνικών επαγωγικής στατιστικής και για την πραγματοποίηση γενικεύσεων.
- Άρα η κατανόηση του δείκτη αυτού είναι αναγκαία.

Υπολογίζεται με τον τύπο:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}}$$

Όπου:

- x_i είναι οι τιμές της κατανομής ($x_1, x_2, x_3, \dots, x_N$)
- \bar{x} είναι ο μέσος όρος των τιμών αυτών.
- N είναι το μέγεθος του δείγματος (πλήθος τιμών).

Στο παράδειγμα με το τεστ μαθηματικών, ας υπολογίσουμε την τυπική απόκλιση των βαθμών που πήραν οι μαθητές του τμήματος Α.

Τμήμα Α		Τμήμα Β	
15	13	15	10
16	15	11	20
17	14	19	15

- Οι τιμές x_i είναι: $x_1=13$, $x_2=14$, $x_3=15$, $x_4=15$, $x_5=16$, $x_6=17$.
- Ο μέσος όρος είναι $\bar{x}=15$. Αντικαθιστούμε τις τιμές αυτές στον τύπο της Τ.Α.

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}} = \sqrt{\frac{(13 - 15)^2 + (14 - 15)^2 + (15 - 15)^2 + (15 - 15)^2 + (16 - 15)^2 + (17 - 15)^2}{6}} = \sqrt{\frac{10}{6}} = \sqrt{1,67} = 1,29$$

- Με τον ίδιο τρόπο βρίσκουμε ότι η τυπική απόκλιση των βαθμών του τμήματος Β ισούται με 3,70 μονάδες.
- **Άρα** το τμήμα Β έχει σχεδόν τριπλάσια τυπική απόκλιση ($s=3,70$ μονάδες) σε σχέση με το τμήμα Α ($s=1,29$ μονάδες).
- **Άρα** το Τμήμα Β είναι πολύ πιο ανομοιογενές σε ό,τι αφορά τις επιδόσεις των μαθητών στα μαθηματικά.

ΣΗΜΕΙΩΣΗ: Η τυπική απόκλιση εκφράζεται στην ίδια κλίμακα μέτρησης με τα δεδομένα που περιγράφει. Εάν μια μεταβλητή μετράει το βάρος σε κιλά, η τυπική απόκλιση των τιμών της θα εκφράζει κι αυτή βάρος σε κιλά, κ.ο.κ.

Άσκηση 1

- Σε δείγμα 40 φοιτητών ο μέσος όρος και η τυπική απόκλιση στο μάθημα της στατιστικής ήταν 5 και 3, αντίστοιχα, ενώ στο μάθημα της μεθοδολογίας ήταν 7 και 3,5, αντίστοιχα.
- ***Σε ποιο μάθημα οι βαθμοί παρουσιάζουν μεγαλύτερη ανομοιογένεια και γιατί;***

Μεγαλύτερη ανομοιογένεια παρουσιάζουν οι βαθμοί στο μάθημα της μεθοδολογίας καθώς η τυπική απόκλιση των βαθμών στο μάθημα αυτό είναι μεγαλύτερη από την αντίστοιχη στο μάθημα της στατιστικής ($3,5 > 3$).

Άσκηση 2

- Μια ομάδα δεδομένων περιλαμβάνει μια τιμή $x_1=1,4$. Η τυπική απόκλιση για αυτά τα δεδομένα είναι $s=0$.
- Μπορείτε να υπολογίσετε τον μέσο όρο των δεδομένων αυτών και να πείτε κάτι για τις υπόλοιπες τιμές της ομάδας;

Εφόσον η τυπική απόκλιση είναι ίση με το 0, συμπεραίνουμε πως οι επιμέρους τιμές της κατανομής έχουν μηδενική απόκλιση από το μέσο όρο. Άρα ταυτίζονται με αυτόν. Εφόσον μία από τις τιμές ισούται με 1,4, τότε και ο μέσος όρος θα ισούται με 1,4 και όλες οι άλλες τιμές θα είναι 1,4.

2. ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ ΠΛΗΘΥΣΜΟΥ

Στην περίπτωση που το δείγμα χρησιμοποιείται για τον υπολογισμό της τυπικής απόκλισης του συνολικού πληθυσμού, εφαρμόζουμε τον ίδιο τύπο με μια μικρή «διόρθωση» στον παρονομαστή (βάζουμε $N-1$ αντί και N). Αυτό μας επιτρέπει μια πιο ασφαλή εκτίμηση της τυπικής απόκλισης του πληθυσμού.

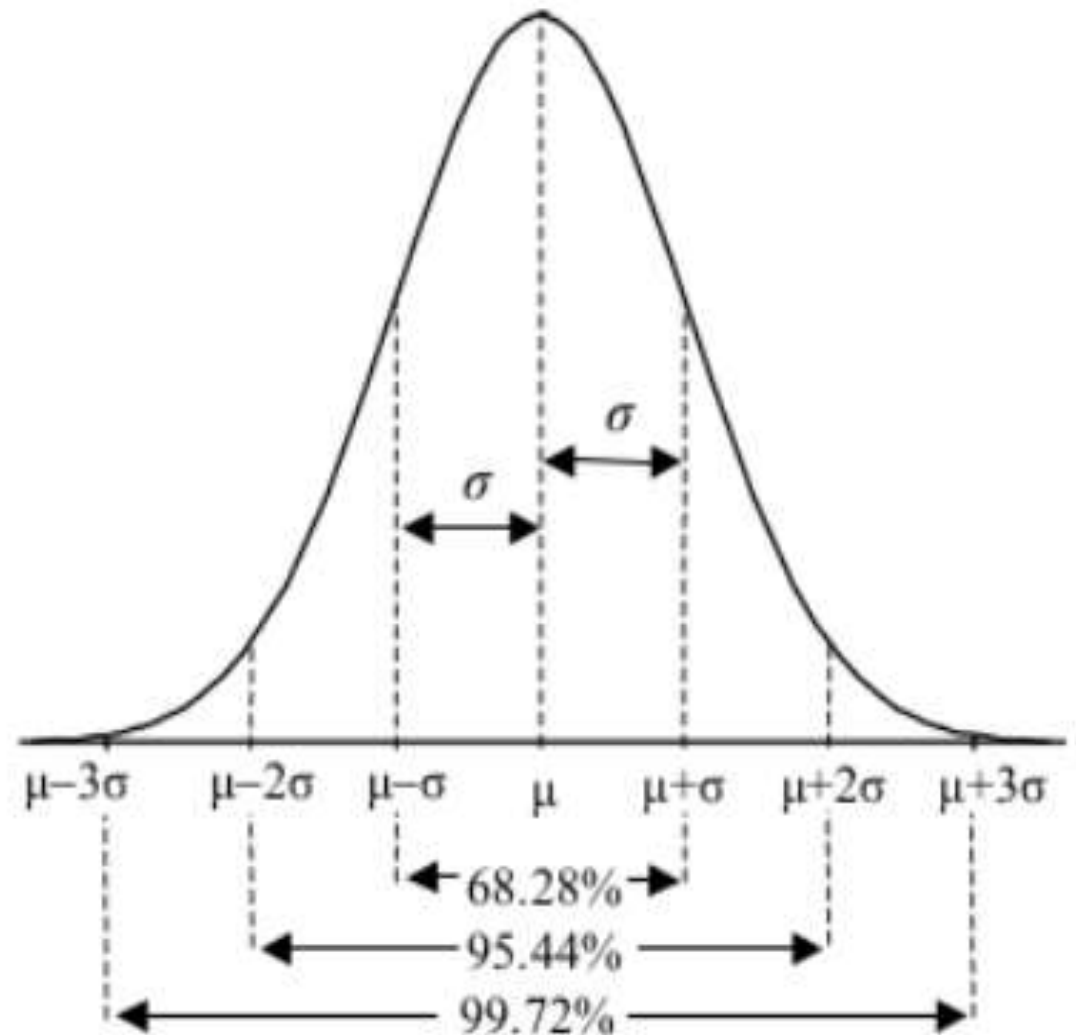
$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N-1}}$$

3. Ο ΚΑΝΟΝΑΣ 68-95-99,7

Εάν ένας πληθυσμός έχει κανονική κατανομή τιμών σε μία μεταβλητή, το ποσοστό των ατόμων που περιλαμβάνονται μεταξύ του μέσου όρου (μ) και:

- ± 1 τυπικής απόκλισης είναι περίπου 68%
- ± 2 τυπικών αποκλίσεων είναι περίπου 95%
- ± 3 τυπικών αποκλίσεων είναι περίπου 99,7%

Ο μέσος όρος (μ) και η τυπική απόκλιση (σ) του πληθυσμού μπορούν να εκτιμηθούν με βάση τον μέσο όρο (\bar{x}) και την τυπική απόκλιση (s) ενός **πιθανοτικού δείγματος** και πάντα με πιθανότητα σφάλματος.



ΒΙΒΛΙΟΓΡΑΦΙΑ

- Mansfield, E. (1983). *Statistics for Business and Economics: Methods and Applications (Second Edition)*. New York: W. W. Norton & Company, Inc.
- Ρούσσος, Π. και Τσαούσης Γ. (2020). *Στατιστική Εφαρμοσμένη στις Κοινωνικές Επιστήμες με τη Χρήση του SPSS και του R*. Αθήνα: Gutenberg.