Μαρίζα Ζυμωνοπούλου marisaz@hmu.gr

4ο Φροντιστήριο – Στατιστική Ι

Κεφ. 6 : Μέτρα κεντρικής τάσης και διασποράς

Ασκήσεις Κατανόησης & Παραδείγματα

Άσκηση 1

Ποιο είναι το μεγαλύτερο πλεονέκτημα του μέσου όρου σε σχέση με τα άλλα δύο μέτρα κεντρικής τάσης, το οποίο μάλιστα τον κάνει να χρησιμοποιείται πιο συχνά στις έρευνες των κοινωνικών επιστημών;

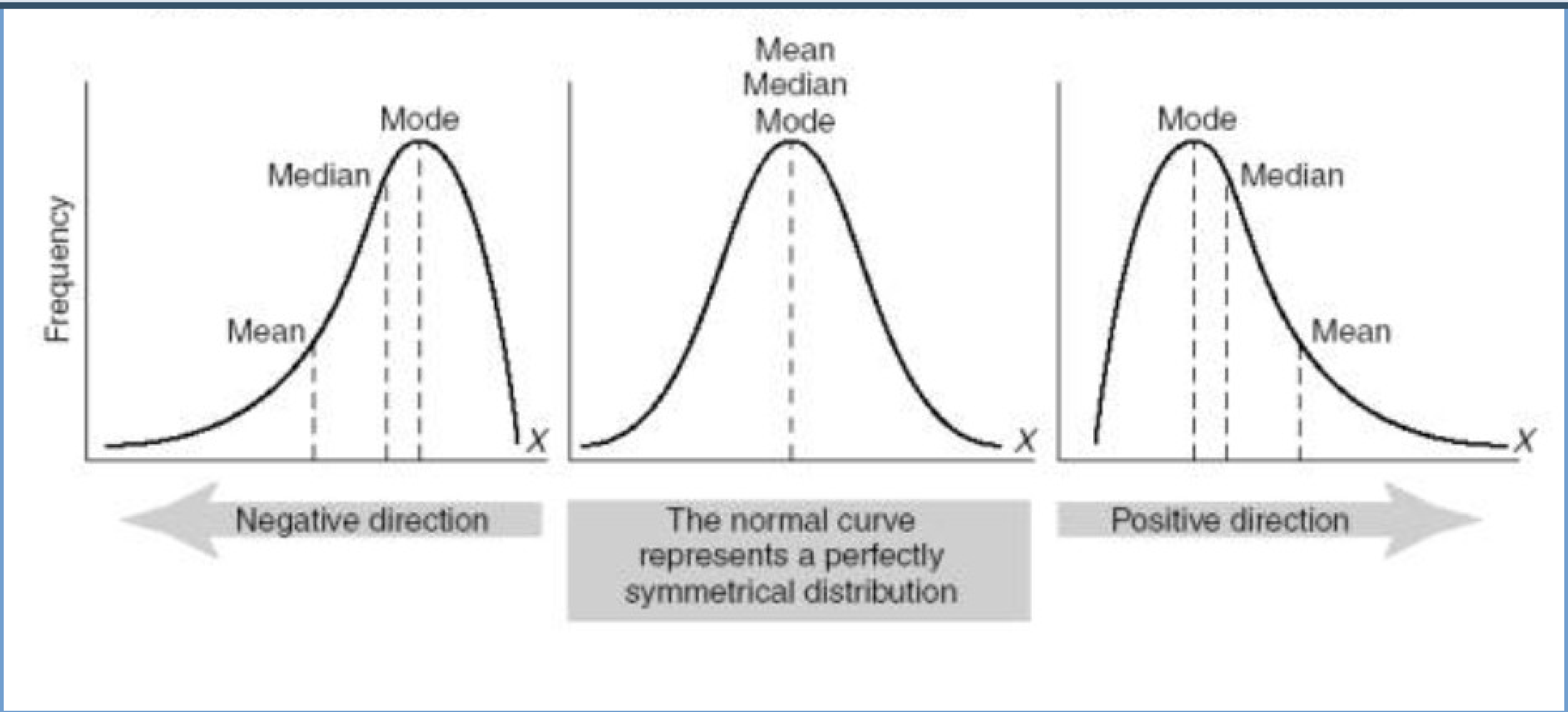
Απάντηση: Το πιο μεγάλο πλεονέκτημα του Μ.Ο. σε σχέση με τα υπόλοιπα μέτρα κεντρικής τάσης είναι ότι μπορεί να χρησιμοποιηθεί και για εξαγωγή συμπερασμάτων για τον πληθυσμό. Χρησιμοποιείται στην επαγωγική στατιστική για τον υπολογισμό παραμέτρων και για τον έλεγχο υποθέσεων (θα δούμε αργότερα τι είναι αυτό...)

Άσκηση 2

Ένας ερευνητής ισχυρίζεται ότι σε μια συμμετρική κατανομή υπολόγισε τα τρία μέτρα κεντρικής τάσης και βρήκε τις παρακάτω τιμές: Δσπ 15, Δμ 18 και Μ.Ο. 16,5. Ένας άλλος ερευνητής θεωρεί ότι αυτές οι τιμές δεν είναι δυνατό να ισχύουν. Γιατί το πιστεύει αυτό;

Απάντηση: Ο δεύτερος ερευνητής έχει δίκιο, αφού όταν η κατανομή είναι κωδωνοειδής συμμετρική (σαν καμπάνα) τότε και τα τρία μέτρα κεντρικής τάσης είναι ίσα.

Ασύμμετρη αριστερά Κωδωνοειδής Συμμετρική Ασύμμετρη δεξιά



Άσκηση 3

Μία ερευνήτρια υπολόγισε τα δύο βασικά μέτρα κεντρικής τάσης των δεδομένων που συγκέντρωσε και βρήκε τις παρακάτω τιμές: Δμ 35 και Μ.Ο. 32,54. Μελετώντας τα αποτελέσματα αυτά είναι σε θέση να έχει μία εκτίμηση για τη μορφή της κατανομής που έχουν τα δεδομένα της;

Απάντηση: Ναι, μπορεί να έχει μία εικόνα. Όταν ο μέσος όρος είναι μικρότερος από τη διάμεσο, η κατανομή είναι ασύμμετρη αριστερά (ανεξάρτητα από την δεσπόζουσα τιμή). Αν ο μέσος όρος ήταν μεγαλύτερος από τη διάμεσο τότε η κατανομή θα ήταν ασύμμετρη δεξιά).

Άσκηση 4

Ας υποθέσουμε ότι τα μέτρα κεντρικής τάσης που παρουσιάζονται στον παρακάτω πίνακα αναφέρονται στις επιδόσεις μαθητών από πέντε διαφορετικά τμήματα στο ίδιο μάθημα. Ποιο τμήμα πιστεύετε ότι τα πήγε καλύτερα; Τεκμηριώστε την άποψή σας.

Άσκηση 4

Τμήματα	1ο	2ο	3ο	4ο	5ο
Μέσος Όρος	15	12	12	14	12
Διάμεσος	17	8	13	13	12
Δεσπόζουσα Τιμή	18	6	7 & 6	13	-

Απάντηση:

Καλύτερα πήγε το 1ο τμήμα γιατί έχει όλα τα μέτρα κεντρικής τάσης υψηλότερα από τα υπόλοιπα τμήματα.

Άσκηση 5

Η Κατερίνα έδωσε εξετάσεις σε τέσσερα μαθήματα το περασμένο εξάμηνο. Ο μέσος όρος των βαθμών της ήταν 6,5, η διάμεσος ήταν 6 και η δεσπόζουσα τιμή 5. Τί βαθμούς έλαβε η Κατερίνα σε κάθε μάθημα;

Έστω ότι οι βαθμοί της Κατερίνας που θέλουμε να υπολογίσουμε είναι $x_1,\ x_2,\ x_3,\ x_4$. Αφού ο μέσος όρος είναι 6,5 παίρνουμε την σχέση:

$$\frac{x_1 + x_2 + x_3 + x_4}{4} = 6,5 \Longrightarrow x_1 + x_2 + x_3 + x_4 = 26 \qquad (1)$$

Επιπλέον, αφού η διάμεσος είναι 6 και έχουμε άρτιο αριθμό τιμών (4), σημαίνει ότι

$$\frac{x_2 + x_3}{2} = 6 \Longrightarrow x_2 + x_3 = 12 \qquad (2)$$

Οπότε

$$x_1+x_2+x_3+x_4 = 26 \Longrightarrow x_1+12+x_4 = 26 \Longrightarrow x_1+x_4 = 26-12 = 14 \qquad (3)$$

Επίσης, εφόσον η δεσπόζουσα τιμή είναι το 5, σημαίνει ότι κάποιοι από τους βαθμούς ήταν 5. Δεν θα μπορούσαν να είναι και οι τέσσερις γιατί ο μέσος όρος είναι 6,5. Οπότε 5 θα είναι είτε δύο τιμές, είτε τρείς. Έτσι, έχουμε τις εξής περιπτώσεις:

1η: $x_1 = x_2 = x_3 = 5$

2η: $x_2 = x_3 = 5$

3η: $x_1 = x_2 = 5$

Οι δύο πρώτες περιπτώσεις δεν μπορούν να ισχύουν, λόγω των σχέσεων (1) και (2). Συνεπώς ισχύει η 3η περίπτωση. Τότε,

$$(2) : x_2 + x_3 = 12 \Rightarrow 5 + x_3 = 12 \Rightarrow x_3 = 7$$

$$(3) : x_1 + x_4 = 14 \Rightarrow 5 + x_4 = 14 \Rightarrow x_4 = 9.$$

Άρα οι βαθμοί είναι: 5, 5, 7, 9

Άσκηση 6

Σε μία τάξη 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. Σε ποιο μάθημα οι βαθμοί των φοιτητών παρουσιάζουν μεγαλύτερη ανομοιογένεια και γιατί;

Απάντηση: Μεγαλύτερη ανομοιογένεια παρουσιάζουν στο μάθημα της Μεθοδολογίας, διότι η τυπική απόκλιση των επιδόσεων είναι μεγαλύτερη από την τυπική απόκλιση των επιδόσεων στη Στατιστική.

Άσκηση 7

Κάποιος σας αναφέρει ότι μία ομάδα δεδομένων περιλαμβάνει μία τιμή, η οποία είναι το 1,4. Η τυπική απόκλιση γι’αυτά τα δεδομένα είναι 0. Μπορείτε να υπολογίσετε τον μέσο όρο των δεδομένων αυτών και να πείτε κάτι για τις υπόλοιπες τιμές της ομάδας;

Απάντηση: Εφόσον η τυπική απόκλιση είναι 0, δεν υπάρχει διακύμανση των τιμών, οπότε όλες οι τιμές είναι ίδιες. Επομένως, εδώ όλες οι τιμές θα είναι 1,4 και τότε ο μέσος όρος θα είναι επίσης 1,4.

Η τυπική απόκλιση (s) ορίζεται ως η ρίζα της διακύμανσης, δηλαδή

$$\text{Διακύμανση: } s^2 = \frac{\sum(x_i - \bar{x})^2}{N} \quad \text{ή} \quad s^2 = \frac{d_i^2}{N}$$

$$\text{Απόκλιση: } s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}} \quad \text{ή} \quad s = \sqrt{\frac{d_i^2}{N}}$$

όπου $d_i = x_i - \bar{x}$

Άσκηση 8

Δημιουργήστε μία μικρή ομάδα δεδομένων περίπου οκτώ τιμών και δείξτε ότι η πρόσθεση ή η αφαίρεση μιας σταθερής τιμής από κάθεμία τιμή δεν πρόκειται να αλλάξει την τιμή της τυπικής απόκλισης. Τι θα συμβεί, ωστόσο, στην τιμή του μέσου όρου όταν προστεθεί ή αφαιρεθεί μια σταθερή τιμή;

Έστω οι τιμές 4, 9, 5, 8, 7, 7, 8, 4. Υπολογίζουμε το μέσο όρο και την τυπική απόκλιση:

$$\text{M.O.} = \frac{4+9+5+8+7+7+8+4}{8} = \frac{52}{8} = 6,5$$

$$s = \sqrt{\frac{(4-6,5)^2+(9-6,5)^2+(5-6,5)^2+(8-6,5)^2+(7-6,5)^2+(7-6,5)^2}{8}+}$$

$$\sqrt{+\frac{(8-6,5)^2+(4-6,5)^2}{8}} = 1,927248$$

Ο υπολογισμός κυρίως της τυπικής απόκλισης έγινε γρήγορα με τη χρήση του R!

Τώρα προσθέτουμε σε κάθε τιμή το 2, οπότε παίρνουμε τις νέες τιμές 6, 11, 7, 10, 9, 9, 10, 6. Υπόλογίζουμε το νέο μέσο όρο και τη νέα τυπική απόκλιση.

$$\text{Νέος Μ.Ο.} = 8,5$$

$$s_N = 1,927248$$

Παρατηρούμε ότι ο μέσος όρος αυξήθηκε κατά 2, ενώ η τυπική απόκλιση δεν άλλαξε!

Άσκηση 9

Στον πίνακα που ακολουθεί παρουσιάζονται οι μέσοι όροι και οι τυπικές αποκλίσεις των βαθμολογιών μαθητών από πέντε διαφορετικά τμήματα στο ίδιο μάθημα. Ποιο τμήμα είχε την καλύτερη επίδοση στο συγκεκριμένο μάθημα και γιατί;

Άσκηση 9

Τμήμα	**1ο**	**2ο**	**3ο**	**4ο**	**5ο**
Μέσος Όρος	11	11	10	14	14
Τυπική Απόκλιση	3	0,3	0,3	2	1

Απάντηση: Το 5ο τμήμα έχει καλύτερη απόδοση διότι έχει υψηλότερο μέσο όρο και μικρότερη απόκλιση (από αυτά που έχουν υψηλό μέσο όρο).

Άσκηση 10

Θέλετε να περιγράψετε αριθμητικά τη διασπορά των τιμών μιας κατανομής και γνωρίζετε ότι ο συνηθέστερος τρόπος μιας τέτοιας περιγραφής είναι μέσω της τυπικής απόκλισης. Ωστόσο, σε ποια περίπτωση δεν θα έπρεπε να την επιλέξετε;

Απάντηση:

Στην περίπτωση που η κατανομή έχει ακραίες τιμές είναι καλύτερο να χρησιμοποιήσουμε το ενδοτεταρτημοριακό εύρος.

Στην περίπτωση που η κλίμακα είναι ιεραρχική προτιμάμε το εύρος.

Παράδειγμα

Μία ερευνήτρια ενδιαφέρεται να μελετήσει την εβδομαδιαία έκθεση νέων στην τηλεόραση. Τα δεδομένα που συνέλεξε από 10 νέους/ες είναι τα εξής: 7, 3, 8, 4, 6, 1, 8, 4 8, 10. Να υπολογίσετε τα κεντρικά μέτρα θέσης και διασποράς, δηλαδή μέσο όρο, διάμεσο, επικρατούσα (δεσπόζουσα) τιμή, διασπορά, εύρος, τυπική απόκλιση, ενδοτεταρτημοριακό εύρος.

<u>Λύση:</u> Βάζουμε τις τιμές σε αύξουσα σειρά

1, 3, 4, 4, 6, 7, 8, 8, 8, 10

Μέσος όρος τιμών

$$\text{M.O.} = \frac{1+3+4+4+6+7+8+8+8+10}{10} = 5,9$$

Διακύμανση

$$s^2 = \frac{(1-5,9)^2+(3-5,9)^2+(4-5,9)^2+(4-5,9)^2+(6-5,9)^2}{10} +$$

$$+\frac{(7-5,9)^2+(8-5,9)^2+(8-5,9)^2+(8-5,9)^2+(10-5,9)^2}{10} = 7,09$$

Τυπική απόκλιση

$$s = \sqrt{s^2} = 2,6627$$

Οι υπολογισμοί μπορούν να γίνουν άμεσα με τη χρήση R.

Επικρατούσα ή δεσπόζουσα τιμή: Είναι η τιμή που εμφανίζεται τις περισσότερες φορές. Άρα 8.

Εύρος: Είναι η μεγαλύτερη μείον την μικρότερη τιμή, δηλ. 10-1=9.

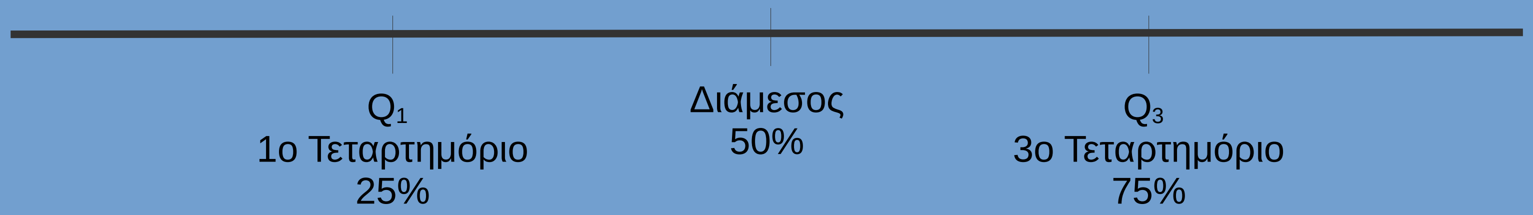
Διάμεσος: Η θέση της διαμέσου εξαρτάται από το δείγμα.

Αν N άρτιος τότε η διάμεσος είναι ανάμεσα στις τιμές $\frac{N}{2}$ και $\frac{N}{2}+1$.

Αν N περιττός τόε η διάμεσος είναι ακριβώς η τιμή που βρίσκεται στη θέση $\frac{N+1}{2}$.

Στην περίπτωσή μας, $N = 10$, οπότε η διάμεσος θα είναι ανάμεσα στον 5° και στον 6° όρο, δηλαδή θα είναι το $\frac{6+7}{2} = 6,5$.

Ενδοτεταρτημοριακό εύρος: είναι το διάστημα στο οποίο συγκεντρώνεται το “κεντρικό” 50% των τιμών. Το ενδοτεταρτημοριακό εύρος καθορίζεται από το Q_1 και το Q_3, όπου Q_1 είναι το σημείο μέχρι το 25% των παρατηρήσεων και Q_3 το σημείο μέχρι το 75% των παρατηρήσεων.



Για τον υπολογισμό του Q_1 θέλουμε να βρούμε την τιμή που βρίσκεται στο 1ο τέταρτο των παρατηρήσεων, δηλαδή στο $\frac{N+1}{4} = \frac{10+1}{4} = 2,75$. Η τιμή αυτή θα βρίσκεται ανάμεσα στον 2° και στον 3° όρο, σε ¨απόσταση' 0,75 ανάμεσά τους. Οπότε θα είναι ο

$$2^{ος} + 0,75(3^{ος} - 2^{ος}) \quad \text{δηλ.} \quad Q_1 = 3 + 0,75(4-3) = 3,75$$

Για τον υπολογισμό του Q_3 θέλουμε να βρούμε την τιμή που βρίσκεται στο 3ο τέταρτο των παρατηρήσεων, δηλαδή στο $\frac{3(N+1)}{4} = \frac{3(10+1)}{4} = 8,25$. Η τιμή αυτή θα βρίσκεται ανάμεσα στον 8° και στον 9° όρο, σε ¨απόσταση' 0,25 ανάμεσά τους. Οπότε θα είναι ο

$$8^{ος} + 0,25(9^{ος} - 8^{ος}) \quad \text{δηλ.} \quad Q_3 = 8 + 0,25(8-8) = 8$$

Συνεπώς το ενδοτεταρτημοριακό εύρος θα είναι

$Q_3 - Q_1 = 8 - 3{,}75 = 4{,}25$

ΤΈΛΟς