ΜΕΤΡΑ ΘΕΣΗΣ ΚΑΙ ΔΙΑΣΠΟΡΑΣ – ΑΡΙΘΜΗΤΙΚΑ ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ

ΜΑΘΗΜΑ: ΠΙΘΑΝΟΤΗΤΕΣ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ

ΤΜΗΜΑ: ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ

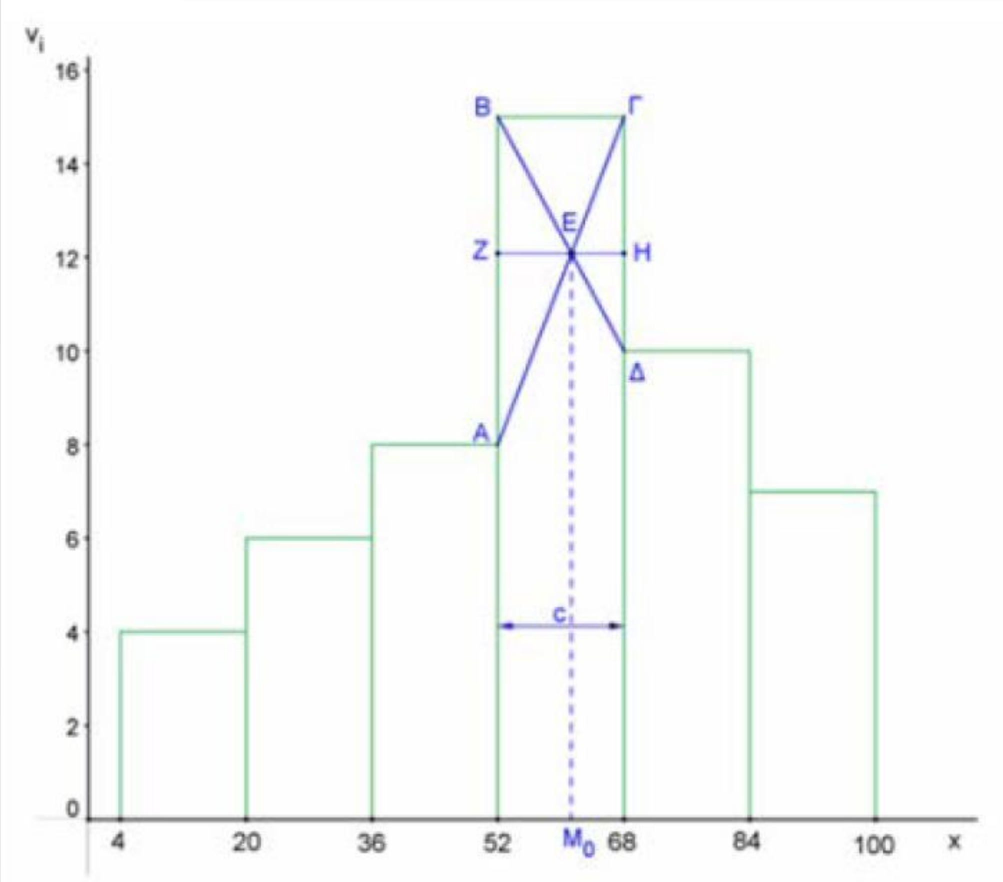
ΚΟΡΥΦΗ Ή ΕΠΙΚΡΑΤΟΥΣΑ ΤΙΜΗ (MODE)

Εκφράζει την τιμή του αποτελέσματος που εμφανίζεται συχνότερα όταν τα δεδομένα μας δεν είναι ομαδοποιημένα.

Παρατήρηση 1: Σε αντίθεση με τα άλλα μέτρα θέσης η επικρατούσα τιμή μπορεί να οριστεί εκτός από ποσοτικά δεδομένα και σε ποιοτικά δεδομένα.

Παρατήρηση 2: Οι κατανομές με δύο προφανείς κορυφές, ακόμα και αν δεν έχουν ακριβώς το ίδιο ύψος, αναφέρονται ως δικόρυφες. *Οι κατανομές με περισσότερες από δύο κορυφές ονομάζονται πολλαπλών κορυφών.*

ΚΟΡΥΦΗ Ή ΕΠΙΚΡΑΤΟΥΣΑ ΤΙΜΗ (MODE)



Με ομαδοποιημένα δεδομένα σε **ισοπλατείς** κλάσεις.
Παράδειγμα (διπλανό σχήμα):
Επικρατούσα τιμή:

Mo = 52 + EZ = 61,33

Από την ομοιότητα των τριγώνων ΑΕΒ και ΓΔΕ υπολογίζουμε ότι το ύψος ΕΖ είναι ίσο με 9,33 λαμβάνοντας υπόψη ότι για δύο όμοια τρίγωνα ο λόγος της ομοιότητάς τους είναι ίσος με το λόγο των δύο ομόλογων υψών τους

Άρα Mo = 52 + EZ = 61,33

ΕΚΑΤΟΣΤΗΜΟΡΙΑ

- Υπολογίζονται για ποσοτικές μεταβλητές σε ένα δείγμα (ή πληθυσμό)

- k-εκατοστημόριο μια μεταβλητής ονομάζουμε την τιμή P_k για την οποία το πολύ k% των παρατηρήσεων του δείγματος (ή του πληθυσμού) είναι μικρότερες από αυτή και το πολύ (1- k)% είναι μεγαλύτερες από αυτή.

- τα P_{25}, P_{50}, P_{75} ονομάζονται τεταρτημόρια και συμβολίζονται με Q_1, Q_2, Q_3

ΕΚΑΤΟΣΤΗΜΟΡΙΑ

Υπολογισμός p-εκατοστημόριου δείγματος:

Διατάσσουμε τις παρατηρήσεις κατά αύξουσα σειρά και υπολογίζουμε το

$$\frac{i(n+1)}{100} = \gamma_i, i = 1,2, \ldots, 99$$

Αν γ_i ακέραιος, τότε το i-εκατοστημόριο είναι η παρατήρηση x_{γ_i}

Διαφορετικά, $P_i = \frac{x_a + x_{a+1}}{2}$ όπου α το ακέραιο μέρος του γ_i

ΕΚΑΤΟΣΤΗΜΟΡΙΑ

Υπολογισμός 1ου, 2ου και 3ου εκατοστημόριου δείγματος:

Διατάσσουμε τις παρατηρήσεις κατά αύξουσα σειρά και υπολογίζουμε το

$$\frac{i(n+1)}{4} = \gamma_i, i = 1,2,3$$

Αν γ_i ακέραιος, τότε το i-εκατοστημόριο είναι: $Q_i = x_{\gamma_i}\ i = 1,2,3$

Διαφορετικά, $Q_i = \frac{x_a + x_{a+1}}{2}$ όπου α το ακέραιο μέρος του γ_i

ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ Ή ΜΕΤΡΑ ΜΕΤΑΒΛΗΤΌΤΗΤΑΣ (MEASURES OF VARIABILITY)

Τα μέτρα αυτά σχετίζονται με την απόκλιση των τιμών από ένα μέτρο θέσης και ονομάζονται ***μέτρα διασποράς***

Τα κυριότερα από τα μέτρα διασποράς είναι:

- το **εύρος R ή κύμανση**
- **η διασπορά ή διακύμανση s^2** και
- **η τυπική απόκλιση s.**

ΕΥΡΟΣ R Ή ΚΥΜΑΝΣΗ

Το **εύρος R ή κύμανση:** Η διαφορά της μικρότερης τιμής από τη μεγαλύτερη τιμή των παρατηρήσεων δηλαδή:
R = (Μεγαλύτερη τιμή των παρατηρήσεων) - (Μικρότερη τιμή των παρατηρήσεων)

Σε ομαδοποιημένο δείγμα παρατηρήσεων το εύρος είναι:

R = (Άνω άκρο τελευταίας κλάσης)-(Κάτω άκρο πρώτης κλάσης).

Ενδοτεταρτημοριακό εύρος ενός δείγματος παρατηρήσεων μια ποσοτικής μεταβλητής:

$$Q = Q_3 - Q_1$$

ΔΙΑΚΥΜΑΝΣΗ Ή ΔΙΑΣΠΟΡΑ (S^2)

Μπορεί να μας δώσει εικόνα για το πώς είναι διάσπαρτες οι παρατηρήσεις γύρω από τη μέση τιμή.

$$\boxed{s^2 = \frac{1}{\nu}\sum_{i=1}^{\nu}(t_i - \bar{x})^2}$$

Σε δείγμα ομαδοποιημένων παρατηρήσεων:

$$\boxed{s^2 = \frac{1}{\nu}\sum_{i=1}^{\nu}(x_i - \bar{x})^2 \nu_i}$$

Όπου x_i είναι οι τιμές (ή τα κέντρα των κλάσεων), v_i οι αντίστοιχες συχνότητες και i=1,2,…,k

Παρατήρηση: Δεν εκφράζεται στις μονάδες που εκφράζονται οι παρατηρήσεις αλλά στο τετράγωνό τους.

ΔΙΑΚΥΜΑΝΣΗ Ή ΔΙΑΣΠΟΡΑ (S^2)

Άλλοι τύποι υπολογισμού:

$$s^2 = \overline{(x^2)} - (\overline{x})^2$$

Σε δείγμα ομαδοποιημένων παρατηρήσεων:

$$\boxed{s^2 = \sum_{i=1}^{\kappa} x_i^2 f_i - \left(\sum_{i=1}^{\kappa} x_i f_i\right)^2}$$

ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ (S)

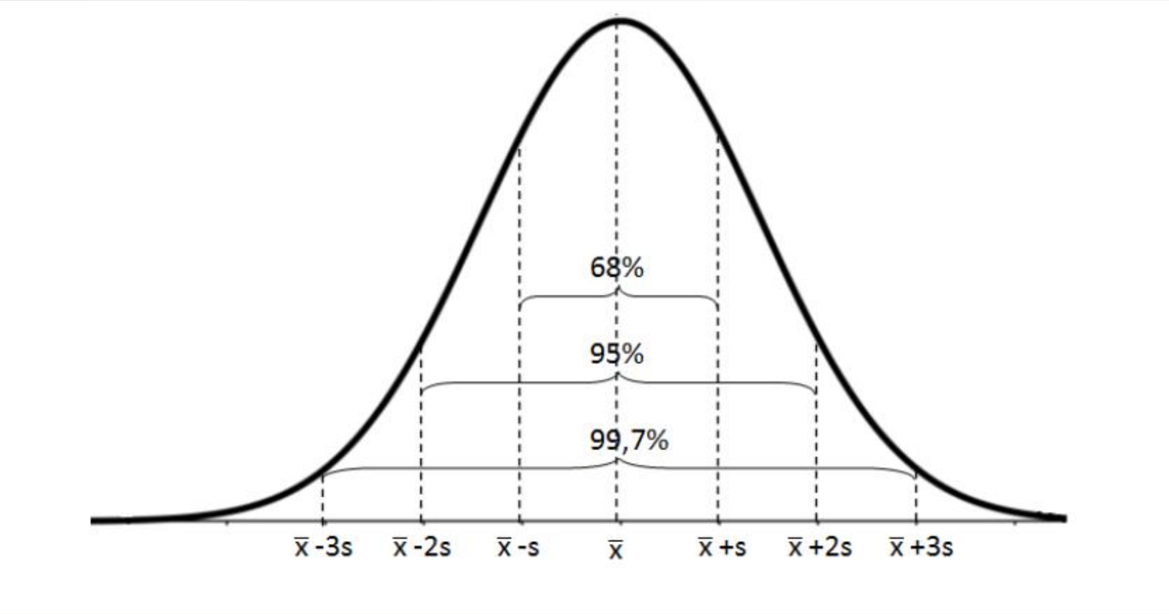
Ορίζεται ως τυπική απόκλιση s η τετραγωνική ρίζα της διακύμανσης δηλαδή:

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

Έχει την ίδια μονάδα μέτρησης με τα μέτρα θέσης

ΠΑΡΑΤΗΡΗΣΕΙΣ:

Όταν το πολύγωνο συχνοτήτων έχει τη μορφή του παρακάτω σχήματος (κανονική κατανομή), τότε η μέση τιμή ταυτίζεται με τη διάμεσο και για την τυπική απόκλιση και αποδεικνύεται ότι:



1) Το 68% των παρατηρήσεων βρίσκεται στο διάστημα:

$$(\bar{x} - s, \bar{x} + s)$$

2) Το 95% των παρατηρήσεων βρίσκεται στο διάστημα: $(\bar{x} - 2s, \bar{x} + 2s)$

3) Το 99.7% των παρατηρήσεων βρίσκεται στο διάστημα: $(\bar{x} - 3s, \bar{x} + 3s)$

4) Το εύρος R είναι περίπου ίσο με έξι τυπικές αποκλίσεις, δηλαδή: $R \approx 6s$

ΠΑΡΑΤΗΡΗΣΕΙΣ

● Αν προσθέσουμε ή αφαιρέσουμε σταθερό αριθμό α από όλες τις τιμές ενός δείγματος, τότε προκύπτει νέο δείγμα με την ίδια τυπική απόκλιση.

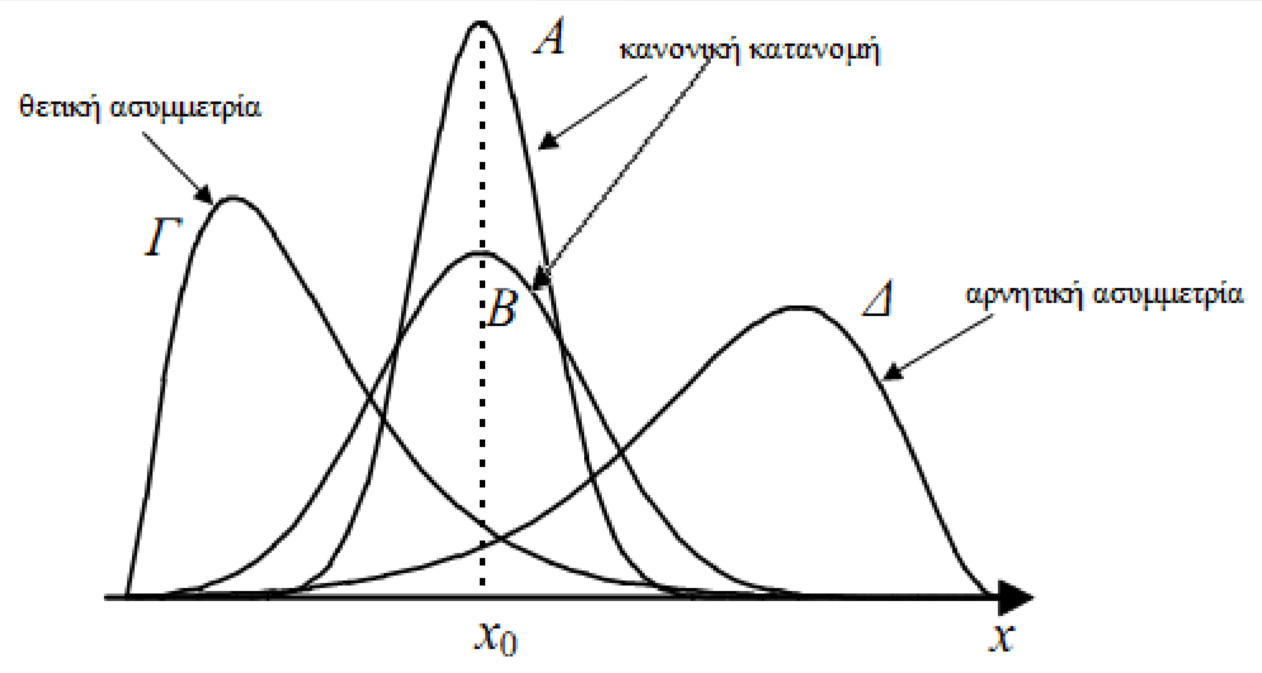
● Αν πολλαπλασιάσουμε με τον ίδιο σταθερό πραγματικό αριθμό λ όλες τις τιμές των παρατηρήσεων ενός δείγματος Α με τυπική απόκλιση s_A, τότε προκύπτει δείγμα Β με τυπική απόκλιση $s_B = |\lambda|\, s_A$

ΜΕΤΡΑ ΑΣΥΜΜΕΤΡΙΑΣ

Καθορίζουν τη μορφή της κατανομής, δηλαδή **κατά πόσο η αντίστοιχη καμπύλη συχνοτήτων είναι συμμετρική ή όχι ως προς την ευθεία $x = x_0$.**

Θετική ασύμμετρη κατανομή: Μια κατανομή που περιλαμβάνει λίγες ακραίες παρατηρήσεις στη θετική κατεύθυνση (δεξιά από την πλειοψηφία των παρατηρήσεων).

Αρνητική ασύμμετρη κατανομή: Μια κατανομή που περιλαμβάνει λίγες ακραίες παρατηρήσεις στη αρνητική κατεύθυνση (αριστερά από την πλειοψηφία των παρατηρήσεων)



ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΤΑΒΛΗΤΗΣ Η΄ ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ

Ο **συντελεστής μεταβολής ή συντελεστής μεταβλητότητας** ενός συνόλου παρατηρήσεων (CV) ποσοτικής μεταβλητής, ορίζεται ως (επί τοις εκατό) το πηλίκο της τυπικής απόκλισης s προς τη μέση τιμή $\bar{x}$ των παρατηρήσεων. Αν είναι $\bar{x} < 0$, τότε χρησιμοποιούμε την $|\bar{x}|$. Δηλαδή:

$$CV = \frac{s}{|\bar{x}|}$$

για κάθε $\bar{x} \neq 0$

Ένα δείγμα τιμών μια μεταβλητής λέγεται **ομοιογενές**, όταν CV ≤ 0,1 ή CV ≤ 10%, ενώ αν CV είναι μεγαλύτερος του 10% λέγεται **ανομοιογενές**.

ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΤΑΒΛΗΤΗΣ Η΄ ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ

- μετρά το «άπλωμα» των δεδομένων σε σχέση με τη μέση τιμή

- εκφράζεται συνήθως ως ποσοστό και είναι καθαρός αριθμός

- μπορεί να χρησιμεύσει για σύγκριση συνόλων δεδομένων που έχουν μετρηθεί με τις ίδιες μονάδες αλλά διαφέρουν σε κάποιο βαθμό ώστε μια άμεση σύγκριση των αντίστοιχων τυπικών αποκλίσεων να μην είναι χρήσιμη.

- χρησιμοποιείται ως μέτρο σύγκρισης της μεταβλητότητας διαφορετικών δειγμάτων με ίδιες ή διαφορετικές μονάδες μέτρησης.

ΒΙΒΛΙΟΓΡΑΦΙΑ:

1. Βιβλίο [86055461]: Στατιστική: Ανάλυση δεδομένων με χρήση της R, Witte Robert, Witte John, Ανδρουλάκης Γεώργιος, Κουνετάς Κωνσταντίνος, Εκδόσεις Κριτική Α.Ε. , 2019.
2. Βιβλίο [77106795]: Στατιστικές μέθοδοι: Θεωρία και εφαρμογές με χρήση Excel και R, Ιωαννίδης Δημήτριος, Εκδόσεις Τζιόλα & Υιοι Α.Ε., 2018.
3. Βιβλίο [77107287]: Βασικές Αρχές Στατιστικής για Επιχειρήσεις-Έννοιες και Εφαρμογές, Berenson L. Mark, Levine M. David, Szabat A. Kathryn, εκδ. Broken Hill Publishers Ltd, 2018.
4. «Μαθηματικά και Στοιχεία Στατιστικής», Γ΄ Γενικού Λυκείου, Ινστιτούτο Τεχνολογίας Υπολογιστών και Εκδόσεων «ΔΙΟΦΑΝΤΟΣ»
5. Complete Business Statistics, by Amir D. Aczel D. & Jayavel Sounderpandian 6th edition
6. Μαθηματικά Γ΄ Λυκείου, Ψηφιακά Εκπαιδευτικά Βοηθήματα, Υπουργείο Παιδείας Έρευνας και Θρησκευμάτων (http://www.study4exams.gr/math_g/index.php)
7. Πιθανότητες και Στατιστική για Μηχανικούς, Νίκος Μυλωνάς, Βασίλης Παπαδόπουλος, Εκδόσεις Τζιόλα