Βασικά Χαρακτηριστικά Αριθμητικών Δεδομένων

- Τάση συγκέντρωσης → Μέτρα Κεντρικής Τάσης και Θέσης

- Τάση διασποράς → Μέτρα Διασποράς

- Σχήμα → Σχήμα της κατανομής

Μέτρα Κεντρικής Τάσης & Θέσης

- **Αριθμητικός Μέσος**
- **Γεωμετρικός Μέσος**
- **Αρμονικός Μέσος**

<u>Μέτρα Κεντρικής Τάσης</u>

- **Διάμεσος ή Κεντρική Τιμή**
- **Τεταρτημόρια**
- **Τύπος ή Επικρατούσα Τιμή**

<u>Μέτρα Κεντρικής Θέσης</u>

Αριθμητικός Μέσος

Απλός ή Αστάθμιστος Αριθμητικός Μέσος

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Παράδειγμα

2	2,1	2,3	3,2	3,2	2,3
3,3	4,1	3,3	3,5	3,5	3,5
3,7	3,7	3,7	5,7	3,9	3,9
3,9	3,9	4,1	4,3	4,3	4,3
4,4	4,4	4,5	5,7	5,7	6

$$\overline{X} = \frac{116{,}4}{30} = 3{,}88$$

Σταθμικός Αριθμητικός Μέσος

$$\overline{X} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

Παράδειγμα

Τιμή (X)	Ποσότητα (w)	$w_i X_i$
120	1000	120000
150	5000	750000
180	6000	1080000
200	3000	600000
	15000	**2550000**

$$\overline{X} = \frac{2550000}{15000} = 170$$

Αριθμητικός Μέσος Κατανομής Συχνοτήτων

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

Παράδειγμα

X_i	f_i	$f_i X_i$
2,0	1,0	2,0
2,1	1,0	2,1
2,3	2,0	4,6
3,2	2,0	6,4
3,3	2,0	6,6
3,5	3,0	10,5
3,7	3,0	11,1
3,9	4,0	15,6
4,1	2,0	8,2
4,3	3,0	12,9
4,4	2,0	8,8
4,5	1,0	4,5
5,7	3,0	17,1
6,0	1,0	6,0

$$\overline{X} = \frac{116{,}4}{30} = 3{,}88$$

Αριθμητικός Μέσος Κατανομής Συχνοτήτων
(Τάξεις μεγέθους)

$$\overline{X} \approx \frac{\sum_{i=1}^{k} f_i X_i'}{\sum_{i=1}^{k} f_i}$$

Παράδειγμα

	f_i	X_i	$f_i X_i$
2,0 - 2,9	4	2.45	9.8
3,0 - 3,9	14	3.45	48.3
4,0 - 4,9	8	4.45	35.6
5,0 - 6,0	4	5.5	22
	30		115.7

$$\overline{X} \approx \frac{115,5}{30} = 3,85$$

- Αποτελεί το πιο σημαντικό μέτρο κεντρικής τάσης

- Επηρεάζεται από όλες τις τιμές

Γεωμετρικός Μέσος

Απλός Γεωμετρικός Μέσος

$$G = \sqrt[n]{\prod_{i=1}^{n} X_i}$$

ή

$$\log G = \frac{1}{n}\sum_{i=1}^{n} \log X_i$$

Παράδειγμα

2,0	2,1	2,3	3,2	3,2	2,3
3,3	4,1	3,3	3,5	3,5	3,5
3,7	3,7	3,7	5,7	3,9	3,9
3,9	3,9	4,1	4,3	4,3	4,3
4,4	4,4	4,5	5,7	5,7	6,0

$$G = e^{\left(\frac{39,62051}{30}\right)} = 3,75$$

Γεωμετρικός Μέσος Κατανομής Συχνοτήτων

$$G = \sqrt[\sum_{i=1}^{k} f_i]{\prod_{i=1}^{k} X_i^{f_i}}$$

ή

$$\log G = \frac{1}{\sum_{i=1}^{k} f_i} \sum_{i=1}^{k} f_i \log X_i$$

Ιδιότητες Γεωμετρικού Μέσου

α) Για να υπολογιστεί απαιτείται $X_i > 0 \quad \forall i$

β) $G \leq \overline{X}$

γ) Επηρεάζεται λιγότερο από τις ακραίες τιμές

δ) Δεν έχει ευρεία εφαρμογή

Αρμονικός Μέσος

Απλός Αρμονικός Μέσος

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

Αρμονικός Μέσος Κατανομής Συχνοτήτων

$$H = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} f_i \frac{1}{X_i}}$$

Ιδιότητες Αρμονικού Μέσου

$$H \le G \le \overline{X}$$

Παράδειγμα

Ένα προϊόν πωλείται σε τρία καταστήματα σε διαφορετική τιμή. Με 10000 δρχ. μπορούν να αγοραστούν 10 μονάδες στο 1ο κατάστημα, στο 2ο κατάστημα 8 μονάδες και στο 3ο 16 μονάδες. Ζητείται ο αριθμός των μονάδων που μπορούν να αγοραστούν κατά μέσο όρο με 10000 δρχ.

Α. Μέσος = (10+8+16)/3 = 11,333 $\Rightarrow$ Μέση Τιμή = 10000/11,333 = 882,3

Πραγματική Μέση Τιμή = (1000+1250+625)/3 = 958,3

Μέσος αριθμός μονάδων = Δαπάνη / Μέση Τιμή

$$\frac{10000}{\dfrac{\frac{10000}{10}+\frac{10000}{8}+\frac{10000}{16}}{3}} = \frac{3}{\frac{1}{10}+\frac{1}{8}+\frac{1}{16}} = 10,43$$

Αρμονικός Μέσος

Διάμεσος ή Κεντρική Τιμή (Median)

Χωρίζει την σειρά των δεδομένων στην μέση αν αυτά τοποθετηθούν κατά αύξουσα ή φθίνουσα τάξη μεγέθους . Είναι η τιμή που βρίσκεται στο μέσο αυτών. Είναι χρήσιμο μέτρο για δείγματα με μεγάλο μέγεθος.

Θέση Διαμέσου $(n+1)/2$

Παράδειγμα

2,0	3,5	4,3
2,1	3,7	4,3
2,3	3,7	4,3
2,3	3,7	4,4
3,2	3,9	4,4
3,2	3,9	4,5
3,3	3,9	5,7
3,3	3,9	5,7
3,5	4,1	5,7
3,5	4,1	6,0

Αν n άρτιος η διάμεσος έχει θεωρητική τιμή

Θέση Διαμέσου (30+1)/2=15,5

Διάμεσος = Μ= 3,9

Μέτρα Σχετικής Τοποθεσίας

- Μέτρα Σχετικής Τοποθεσίας σχεδιάζονται για να προβάλουν πληροφόρηση σχετικά με την ***τοποθεσία*** κάποιων συγκεκριμένων τιμών ***σε σχέση*** με ολόκληρο το σύνολο των δεδομένων.

- ***Ποσοστημόριο:*** το *P* ποσοστημόριο είναι η τιμή από την οποία *P* ποσοστό των τιμών είναι *μικρότερο από* την τιμή αυτή και (100-*P*)% είναι μεγαλύτερο από την τιμή αυτή.

- Υποθέστε ότι ο βαθμός του πτυχίου σας είναι το 60^{o} <u>ποσοστημόριο</u> στο έτος σας , το οποίο σημαίνει ότι το 60% των άλλων σκορ ήταν *κάτω* από το δικό σας, ενώ το 40% των άλλων σκορ ήταν *κάτω* από το δικό σας.

Ποσοστημόριο …

- Έχουμε ειδικά ονόματα για το 25ο, 50ο, και 75ο ποσοστημόριο, χαρακτηριστικά ***τεταρτημόρια.***
- Το πρώτο τεταρτημόριο χαρακτηρίζει Q_1 = 25ο ποσοστημόριο.
- Το δεύτερο τεταρτημόριο, Q_2 = 50ο ποσοστημόριο (το οποίο είναι επίσης η διάμεσος).
- Το τρίτο τεταρτημόριο, Q_3 = 75ο ποσοστημόριο.
- Μπορούμε επίσης να αντιστοιχίσουμε ποσοστημόρια σε πεμπτημόρια (quintiles, fifths) και δεκατημόρια (deciles, tenths).

Χρήσιμα Ποσοστημόρια …

- Πρώτο δεκατημόριο = $10^{ο}$ ποσοστημόριο
- Πρώτο τεταρτημόριο, Q_1, = $25^{ο}$ ποσοστημόριο
- Διάμεσος, Q_2, = $50^{ο}$ ποσοστημόριο
- Τρίτο τεταρτημόριο, Q_3, = $75^{ο}$ ποσοστημόριο
- Ένατο δεκατημόριο = $90^{ο}$ ποσοστημόριο

- **Σημειώστε:** Εάν ο βαθμός σου σε φέρνει στο $80^{ο}$ ποσοστημόριο, αυτό δεν σημαίνει ότι απάντησες το 80% των ερωτήσεων σωστά – αυτό σημαίνει ότι το 80% των συμφοιτητών σου είχε σκορ **χαμηλότερο** από το δικό σου. Δείχνει την θέση σου σε σχέση με τους άλλους.

Θέση των Ποσοστημορίων …

- Ο ακόλουθος τύπος μας επιτρέπει να προσεγγίσουμε την θέση του κάθε ποσοστημορίου:

$$L_p = (n+1)\frac{P}{100}$$

Όπου L_p είναι η θέση του P^o ποσοστημόριου

Θέση των Ποσοστημορίων …

- Θεωρήστε τα δεδομένα:
- 0 0 5 7 8 9 12 14 22 33
- Ποια είναι η θέση του 25ο ποσοστημορίου. Δηλαδή, σε ποιο σημείο είναι το 25% των τιμών μικρότερες και 75% των τιμών μεγαλύτερες;

0 0 5 7 8 9 12 14 22 33

- $L_{25} = (10+1)(25/100) = \boxed{2.75}$

Το 25ο ποσοστημόριο είναι τρία-τέταρτα της απόστασης μεταξύ της δεύτερης (που είναι 0) και της τρίτης (που είναι 5) παρατήρησης. Τα τρία-τέταρτα της απόστασης είναι: (.75)(5 – 0) = 3.75
Επειδή η δεύτερη παρατήρηση είναι 0, το 25ο ποσοστημόριο είναι 0 + 3.75 = **3.75**

Θέση των Ποσοστημορίων ...

- Ποιο είναι το τρίτο τεταρτημόριο;

- $L_{75} = (10+1)(75/100) = \boxed{8.25}$

0 0 5 7 8 9 12 14 22 33

Τοποθετείτε ένα-τέταρτο της απόστασης ανάμεσα στην όγδοη και ένατη παρατήρηση, οι οποίες είναι 14 και 22, αντίστοιχα. Το πρώτο τέταρτο της απόστασης είναι: (.25)(22 - 14) = 2, το οποίο σημαίνει ότι το 75^{ο} ποσοστημόριο είναι: 14 + 2 = **16**

Θέση των Ποσοστημορίων …

- Παρακαλώ θυμηθείτε …

θέση 2.75

16

0 0 | 5 7 8 9 12 14 | 22 33

3.75

θέση 8.25

L_p καθορίζει την **θέση** στο σύνολο των δεδομένων όπου η τιμή του ποσοστημορίου βρίσκεται, όχι την τιμή του ποσοστημορίου.

Διάμεσος κατανομής με διαστήματα τάξεων

$$M \approx F_i + \frac{\delta}{f_i}\left(\frac{n}{2} - \Phi_{i-1}\right)$$

F_i = το κατώτερο όριο της κλάσης που περιέχει την διάμεσο

δ = το εύρος της κλάσεως

f_i = συχνότητα της κλάσεως

Φ_{i-1} = αθροιστική συχνότητα προηγούμενης κλάσης

Παράδειγμα

	f_i	X_i	Φ_i
2,0 - 2,9	4	2,45	4
3,0 - 3,9	14	3,45	18
4,0 - 4,9	8	4,45	26
5,0 - 6,0	4	5,45	30
	30		

$$M \approx 3{,}0 + \frac{0{,}9}{14}\left(\frac{30}{2} - 4\right) = 3{,}7$$

Τεταρτημόρια

Χωρίζουν την σειρά των δεδομένων σε 4 ίσα μέρη αν αυτά τοποθετηθούν κατά αύξουσα τάξη μεγέθους

Θέση 1ου Τεταρτημόριου $(n+1)/4$

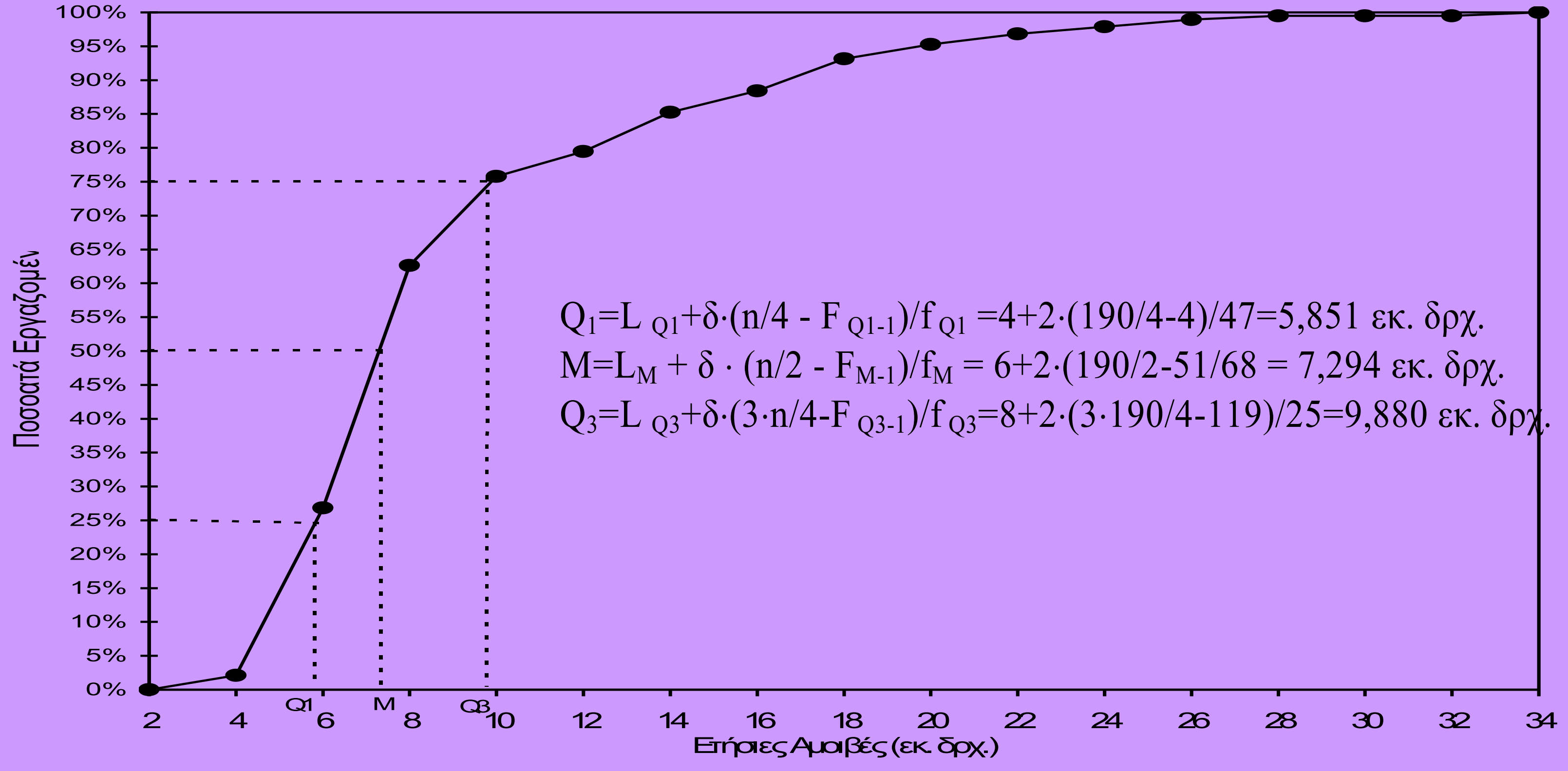
Θέση 2ου Τεταρτημόριου $(n+1)/2$ **(Διάμεσος)**

Θέση 3ου Τεταρτημόριου $3(n+1)/4$

Τύπος ή Επικρατούσα Τιμή (Mode)

Η τιμή με την μεγαλύτερη συχνότητα ν που παρουσιάζεται τις περισσότερες φορές

Εκτίμηση της Διαμέσου και των Τεταρτημορίων με Γραμμική Παρεμβολή



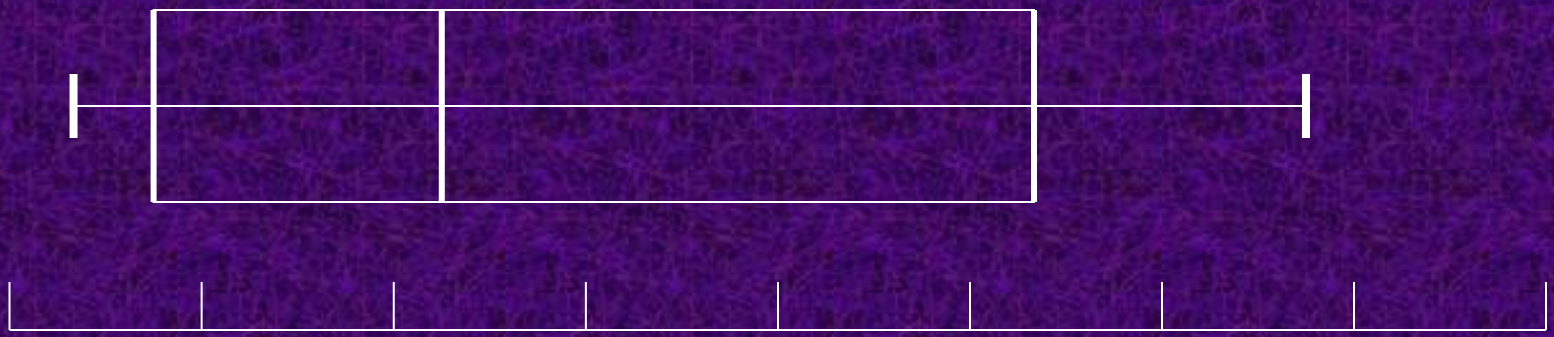
$Q_1=L_{Q1}+\delta\cdot(n/4 - F_{Q1-1})/f_{Q1} =4+2\cdot(190/4-4)/47=5{,}851$ εκ. δρχ.

$M=L_M + \delta \cdot (n/2 - F_{M-1})/f_M = 6+2\cdot(190/2-51/68 = 7{,}294$ εκ. δρχ.

$Q_3=L_{Q3}+\delta\cdot(3\cdot n/4-F_{Q3-1})/f_{Q3}=8+2\cdot(3\cdot 190/4-119)/25=9{,}880$ εκ. δρχ.

Θηκόγραμμα (Box Plot)…

- Το ***θηκόγραμμα*** (***box plot)*** είναι μία γραφική τεχνική η οποία σχεδιάζει **πέντε** στατιστικές:
- την μικρότερη την μεγαλύτερη παρατήρηση, το πρώτο, δεύτερο, και τρίτο τεταρτημόριο.

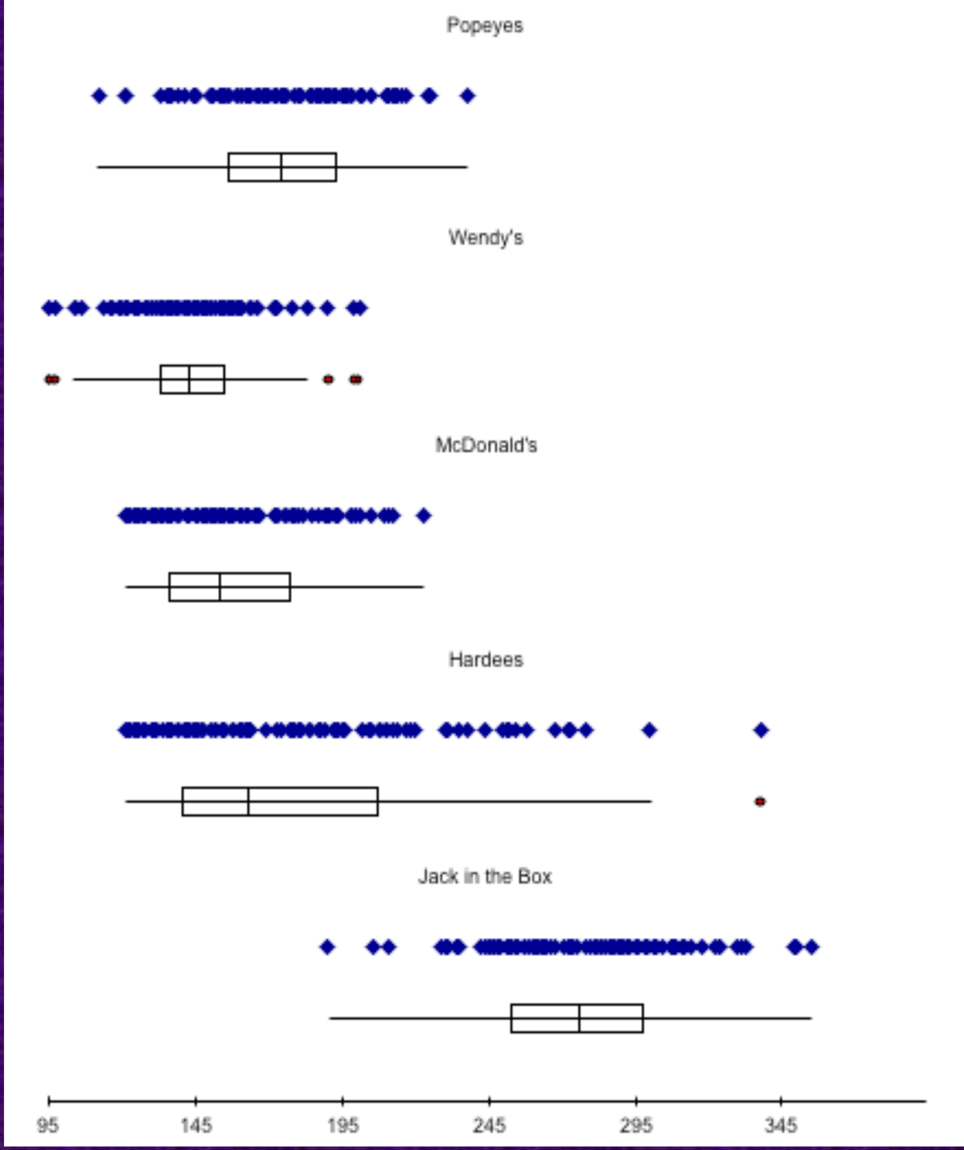


Μύστακας (1.5*(Q_3–Q_1))

Οι προεκτεινόμενες γραμμές στα αριστερά και στα δεξιά καλούνται **μύστακες** (**whiskers**). Κάθε σημείο που πέφτει έξω από τους μύστακες καλείται ακραία τιμή (outlier). Οι μύστακες επεκτείνονται προς τα έξω στο μικρότερο από το (1.5 φορά το ενδοτεταρτημοριακό εύρος) ή στην ποιο τελευταία τιμή η οποία δεν είναι ακραία.

Θηκόγραμμα (Box Plot)…

- Αυτά τα θηκογράμματα είναι από δεδομένα με χρόνος εξυπηρέτησης πελατών.
- Ο χρόνος εξυπηρέτησης στα Wendy’s είναι ο μικρότερος και λιγότερο μεταβλητός.
- Στα Hardee’s ο χρόνος εξυπηρέτησης έχει την μεγαλύτερη μεταβλητότητα, ενώ στα Jack-in-the-Box έχει τον μεγαλύτερο χρόνος εξυπηρέτησης.



Σύγκριση Μέτρων Κεντρικής Τάσης και Θέσης

- Μόνο στα μέτρα Κ. Τάσης μπορούν να γίνουν αλγεβρικές πράξεις

- Ο Αριθμητικός μέσος υπολογίζεται σε κάθε περίπτωση.
  Ο Γεωμετρικός και Αρμονικός Μέσος δεν δέχονται τιμές < = 0

- $\overline{X} \geq G \geq H$

- $\overline{X} = M$ Όταν αναφερόμαστε σε ολόκληρο τον πληθυσμό. Επίσης ισχύει για τις μεταβλητές που ακολουθούν συμμετρικές κατανομές (π.χ. Κανονική, t-κατανομή, Τυπική κατανομή)

- $M < \overline{X}$ Στις κατανομές με σχήμα καμπύλης, κοίλης προς τα δεξιά
- $M > \overline{X}$ Στις κατανομές με σχήμα καμπύλης, κοίλης προς τα αριστερά