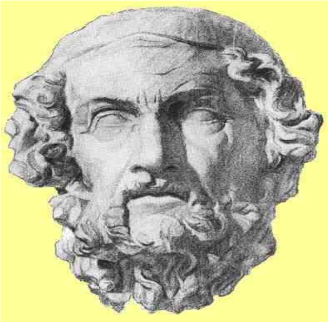
Εισαγωγή στη στατιστική.

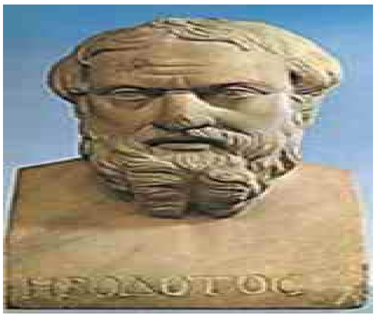
Στατιστική είναι η επιστήμη που σκοπό έχει να εξάγει συμπεράσματα για την εξυπηρέτηση διαφόρων σκοπών, με βάση παρατηρήσεις που έχουν γίνει σχετικά με το προς μελέτη θέμα. Κύριο αντικείμενο της είναι η **συλλογή, παρουσίαση, ανάλυση** στοιχείων ενός μικρού υποσυνόλου δεδομένων και η διατύπωση, με χρήση επαγωγής, των συμπερασμάτων που προκύπτουν από την ανάλυση.

Βρίσκει εφαρμογές σε επιστήμες όπως η φυσική, η χημεία, η μετεωρολογία, η αστρονομία, η κοινωνιολογία[1], η οικονομία, η υγιεινή, η ιατρική, η βιολογία, η ανθρωπολογία, η πολιτική και η τεχνολογία. Η χρησιμοποίηση του όρου πιθανολογείται ότι προέρχεται από τη λατινική λέξη status που σημαίνει κράτος, διότι αρχικά δήλωνε την καταγραφή και κατάταξη ορισμένων γεγονότων που είχαν σημασία για όλο το κοινωνικό σύνολο (γεννήσεις, θάνατοι, γάμοι, ύψος παραγωγής ανά προϊόν, εγκλήματα). Με την έννοια αυτή, η στατιστική συναντάται στις περισσότερες αρχαίες κοινωνίες.

Αναφέρονται περιπτώσεις απογραφής του πληθυσμού και συλλογής στατιστικών στοιχείων για τη ζωή του, στην αρχαία Αθήνα, Σπάρτη, Ρώμη, Αίγυπτο, Κίνα. Τον όρο στατιστική αναφέρουν ο Αριστοτέλης στην «Πολιτεία» και ο Σωκράτης (Ξενοφώντος «Απομνημονεύματα»). Πρώτη ιστορική συλλογή καθαρά στατιστικών στοιχείων, θεωρείται η απογραφή πληθυσμού από τον Αυτοκράτορα της Κίνας Γιάο (Yao) το **2238 πΧ**. Μια μορφή στατιστικής με αριθμητικά δεδομένα, από τον Όμηρο, είναι ο κατάλογος πλοίων των Αχαιών στον Τρωικό πόλεμο, από όπου αντλήθηκαν στοιχεία της οικονομικής ευρωστίας και του πληθυσμού των πόλεων–κρατών που συμμετείχαν και πληροφορίες για την τότε ναυπηγική, ναυτιλία και ναυτική τέχνη.



Εικόνα 1.
Ο Όμηρος.



Εικόνα 2.
Ο Ηρόδοτος.

Ο πατέρας της Ιστορίας Ηρόδοτος (484–420 πΧ), θεωρείται πρωτοπόρος στη στατιστική[2]. Αναφέρει ότι σε συνέλευση του Ξέρξη με τους στρατηγούς του, ο θείος του **Αρτάβανος**, είπε: «Βασιλιά, αν δε λεχθούν όλες οι γνώμες δε μπορεί κανείς να διαλέξει την καλύτερη, αλλά πρέπει να δεχθεί την πρώτη. Διότι βλέποντας το μόνο του, δε ξεχωρίζομε το καθαρό χρυσάφι. Αφότου όμως το τρίψομε στην πέτρα και το συγκρίνομε με άλλο χρυσάφι, τότε αντιλαμβανόμαστε το ποιόν του». Η μέθοδος αυτή

[1] Η πλειονότητα των κατόχων βραβείου Nobel είναι πρωτότοκα παιδιά. Το 65% των μεσαίων παιδιών βάζουν χρήματα στην άκρη. Τα στερνοπούλια επισκέπτονται τα νοσοκομεία λόγω τραυματισμών κατά 50% συχνότερα από τα μεγαλύτερα αδέλφια τους, σύμφωνα με έρευνα στα νοσοκομεία της Ιερουσαλήμ το 2013.

[2] Σύμφωνα με τη μελέτη «Herodotus Statistics – Quantitative commentary» (Τα στατιστικά του Ηροδότου– Ποσοτικά σχόλια), του Βασιλείου Χομπά, Καθηγητή Οικονομικού Τμήματος Παν/μίου Αθηνών, που δημοσιεύθηκε στο αμερικανικό περιοδικό «Journal of the History of Economic Thought».

(της παραγοντικής αναλύσεως) εφαρμόζεται σήμερα κυρίως για την ταξινόμηση απόψεων κατά τις συνεδριάσεις του ΟΗΕ.

Από τις «Ιστορίες» του Ηροδότου, προκύπτει ότι οι Έλληνες ήταν εξοικειωμένοι με τις έννοιες της τύχης, του δυνατού συμβάντος και του αδυνάτου.

Χαρακτηριστικά αναφέρει ο Ηρόδοτος ότι ο βασιλιάς της Αιγύπτου **Άμασις** απευθυνόμενος στον τύραννο της Σάμου Πολυκράτη, λέει: «Είναι χαρά να ακούει κανείς ότι ένας φίλος είναι ευτυχής, αλλά οι μεγάλες επιτυχίες σου δε μου αρέσουν, διότι ξέρω πολύ καλά πόσο φθονερός είναι ο Θεός. Για αυτό και εύχομαι και για εμένα (και για όσους νοιάζομαι), να έχω επιτυχίες αλλά και αποτυχίες και έτσι να ζήσω με τέτοιες εναλλαγές, παρά να έχω πάντα επιτυχίες, διότι για κανέναν από όσους είχαν μόνο επιτυχίες δεν άκουσα ότι δεν πέθανε δυστυχισμένος και κατεστραμμένος». Η θέση αυτή εκφράζεται στη στατιστική με τις δοκιμές Bernoulli, δηλαδή επιτυχίες και αποτυχίες με κάποια πιθανότητα.

Δεν αναφέρονται απογραφές κατά τη διάρκεια του Μεσαίωνα, εκτός από το Liber consualis (στατιστική περιγραφή των μεταλλείων, ιχθυοτροφείων, αγροτεμαχίων) στην Αγγλία του Γουλιέλμου του Κατακτητή (1086 μΧ).

Από τα τέλη του $13^{ου}$ και ιδίως τον $14^{ο}$ αιώνα οι απογραφές αποκτούν πιο συστηματική μορφή στη Βενετία. Στην Αγγλία τον $17^{ο}$ αιώνα αναπτύσσετε η στατιστική χάρη στις εργασίες των Edmund Halley και Sir William Petty (26/05/1623–16/12/1687).

Εικόνα 3.
Ο Edmund Halley.

Εικόνα 4.
Ο Sir William Petty.

Η πρώτη γνωστή δημοσκόπηση στη σύγχρονη ιστορία πραγματοποιήθηκε στην Πενσυλβανία των ΗΠΑ το **1824**. Πεντακόσιοι πολίτες κλήθηκαν να απαντήσουν αν στις επερχόμενες εκλογές θα ψήφιζαν τον **Andrew Jackson**[3] ή τον **John Quincy Adams**[4]. Ο Jackson κέρδισε στη δημοσκόπηση και στις εκλογές.

Μετά από περίπου ενενήντα έτη έγινε η πρώτη παναμερικανική δημοσκόπηση που ανέδειξε νικητή τον **Thomas Woodrow Wilson**[5]. Η στατιστική γνώρισε μεγάλη ανάπτυξη μετά τον $2^{ο}$ Παγκόσμιο πόλεμο λόγω της αναπτύξεως των ηλεκτρονικών υπολογιστών που μπορούσαν να χειρίζονται μεγάλο αριθμό δεδομένων, σε σύντομο χρονικό διάστημα.

---

[3] Ο Andrew Jackson (15/03/1767–08/06/1845) ήταν ο $7^{ος}$ πρόεδρος των ΗΠΑ (04/03/1829 – 04/03/1837).

[4] Ο John Quincy Adams (11/07/1767–23/02/1848) ήταν ο $6^{ος}$ Πρόεδρος των ΗΠΑ (04/03/1825–04/03/1829) και γιος του John Adams, $2^{ου}$ Προέδρου των ΗΠΑ (04/03/1797–04/03/1801).

[5] Ο Thomas Woodrow Wilson (28/12/1856–03/02/1924) εκλέχθηκε $28^{ος}$ Πρόεδρος των ΗΠΑ (04/03/1913–04/03/1921).

Εικόνα 5.
Ο Andrew Jackson.

Εικόνα 6.
Ο John Quincy Adams.

Εικόνα 7.
Ο Thomas Woodrow Wilson.

Εικόνα 8.
Ο George Horace Gallup.

Στην Ευρώπη, οι δημοσκοπήσεις ήρθαν στο τέλος του 2^{ου} Παγκοσμίου πολέμου, πρώτα στην Αγγλία από τον **George Horace Gallup**[6] (18/11/1901–26/07/1984) ο οποίος προέβλεψε την ήττα του έως τότε πρωθυπουργού **Sir Winston Leonard Spencer-Churchill**. Ακολούθως οι δημοσκοπήσεις χρησιμοποιήθηκαν στην κατεχόμενη Γερμανία, από τις συμμαχικές δυνάμεις, για να παρακολουθήσουν τις διαθέσεις της κοινής γνώμης απέναντι στο ναζισμό.

Στην Ελλάδα η πρώτη δημοσκόπηση πραγματοποιήθηκε το **1946** από τη «Συμμαχική επιτροπή για την παρακολούθηση των ελληνικών εκλογών»[7]. Το **1956** έγινε η δεύτερη δημοσκόπηση με αντικείμενο έρευνας τον αντιαμερικανισμό. Στις εκλογές του **1977** παρατηρείτε ραγδαία αύξηση του αριθμού των δημοσκοπήσεων που συνοδεύεται από επιφυλάξεις ως προς την αποτελεσματικότητα τους.

Η στατιστική χωρίζεται σε τρεις κλάδους:

- Το σχεδιασμό πειραμάτων που ασχολείται με τη διαδικασία συλλογής δεδομένων.
- Την περιγραφική στατιστική (descriptive statistics), που ασχολείται με τη συνοπτική και αποτελεσματική παρουσίαση δεδομένων.
- Τη στατιστική συμπερασματολογία ή επαγωγική στατιστική, που ασχολείται με την ανάλυση και εξαγωγή αντίστοιχων συμπερασμάτων.

Η στατιστική μελέτη περιλαμβάνει τις ακόλουθες πέντε φάσεις:

[6] Ο Αμερικανός George Horace Gallup (18/11/1901–26/07/1984) ήταν πρωτοπόρος των τεχνικών δειγματοληψίας και επινόησε τη δημοσκόπηση για τη μέτρηση της κοινής γνώμης.

[7] Βλέπε «Η ανάπτυξη των πολιτικών δημοσκοπήσεων στην Ελλάδα» του Ηλία Νικολακόπουλου, Καθηγητή τμήματος Πολιτικής Επιστήμης & Δημόσιας Διοίκησης, Πανεπιστημίου Αθηνών.

1η φάση	Σχεδιασμός της μελέτης (experimental design)
2η φάση	Συλλογή των στατιστικών στοιχείων
3η φάση	Παρουσίαση των στατιστικών στοιχείων (descriptive statistics)
4η φάση	Ανάλυση των στατιστικών στοιχείων με ειδικές μεθόδους
5η φάση	Εξαγωγή συμπερασμάτων (inferential statistics)

Οι πλέον χαρακτηριστικές έννοιες στην επιστήμη της στατιστικής είναι ο πληθυσμός, το δείγμα και η μεταβλητή.

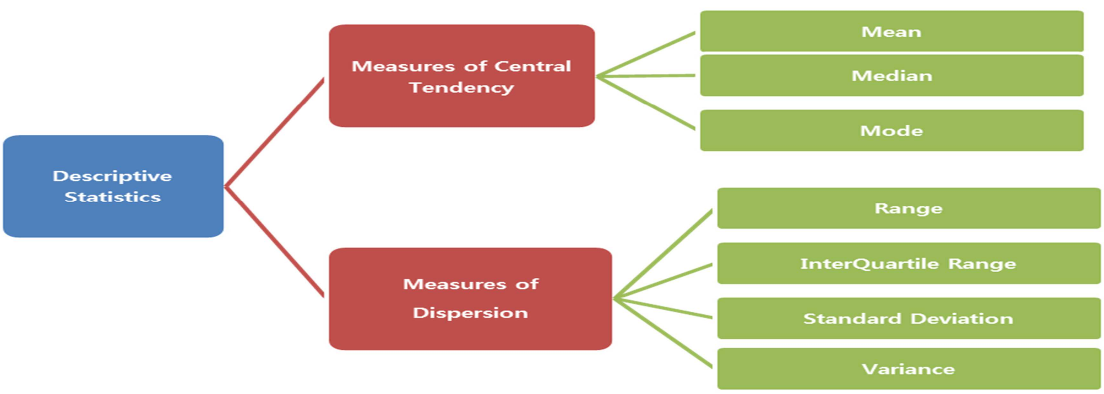
Πληθυσμός, στη στατιστική, ονομάζεται το σύνολο των ανθρώπων ή αντικειμένων που εξετάζονται ως προς ένα ή περισσότερα χαρακτηριστικά τους. (Π.χ. οι οικογένειες μίας πόλεως, οι σπουδαστές μίας σχολής, τα πλοία μίας ναυτιλιακής εταιρείας). Κάθε στοιχείο του πληθυσμού ονομάζεται άτομο. Το πλήθος των ατόμων του πληθυσμού ονομάζεται μέγεθος του πληθυσμού και συμβολίζεται με ν.

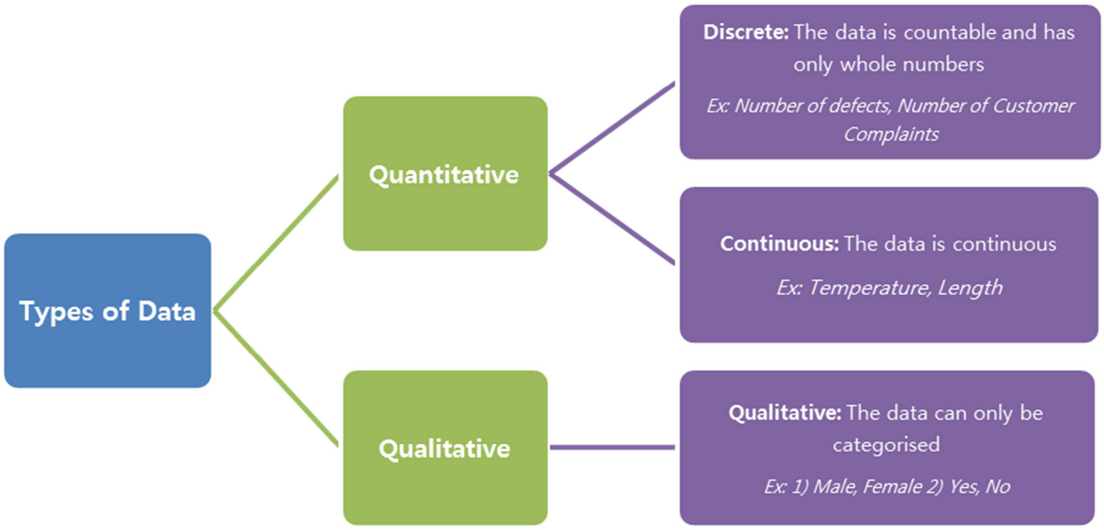
Δείγμα ενός πληθυσμού ονομάζεται μία μικρή ομάδα ανθρώπων ή αντικειμένων που επιλέγονται από τον πληθυσμό. Συνεπώς, το δείγμα είναι υποσύνολο του πληθυσμού. Προκειμένου να εξαχθούν έγκυρα συμπεράσματα, πρέπει το δείγμα να είναι αντιπροσωπευτικό, δηλαδή να έχει επιλεγεί με τέτοιο τρόπο ώστε κάθε μέλος του πληθυσμού να έχει την ίδια πιθανότητα να επιλεγεί. Η εξέταση ενός δείγματος του πληθυσμού, ονομάζεται **δειγματοληψία**.

Μεταβλητή στη στατιστική ονομάζεται το χαρακτηριστικό εκείνο ως προς το οποίο εξετάζεται ο πληθυσμός. Συμβολίζεται με Χ ή Υ ή Ζ, κλπ. Τιμές της μεταβλητής, ονομάζονται οι τιμές εκείνες που μπορεί να λάβει μία μεταβλητή.
Οι μεταβλητές διακρίνονται σε ποιοτικές και ποσοτικές.

Ποιοτικές ονομάζονται οι μεταβλητές των οποίων οι τιμές δεν είναι αριθμοί, δηλαδή δεν επιδέχονται μέτρηση (οικογενειακή κατάσταση, τόπος γεννήσεως, θρησκεία, χρώμα ματιών ή μαλλιών ή αυτοκινήτου, είδος κατοικίας, μάρκα αυτοκινήτου, φύλλο ατόμου, τόπος διακοπών, επάγγελμα, τύπος πλοίων, επίπεδο μορφώσεως). Τα ποιοτικά χαρακτηριστικά πρέπει να είναι αντικειμενικά. Οι χαρακτηρισμοί «ψηλός», «αδύνατος», «καλός», «ωραίος», δεν αποτελούν ποιοτικά χαρακτηριστικά, διότι είναι εντελώς υποκειμενικά.

Ποσοτικές ονομάζονται οι μεταβλητές των οποίων οι τιμές είναι αριθμοί. Οι ποσοτικές μεταβλητές διακρίνονται σε διακριτές και συνεχείς. Οι διακριτές μεταβλητές λαμβάνουν μόνο μεμονωμένες (διακεκριμένες) τιμές (π.χ. πλήθος παιδιών, νούμερο πλοίων, αριθμός δωματίων, ενδείξεις ζαριού).





Οι **συνεχείς** μεταβλητές λαμβάνουν οποιαδήποτε τιμή εντός ενός διαστήματος πραγματικών αριθμών ή ενώσεως διαστημάτων. (π.χ. ύψος ή βάρος σπουδαστών, μισθός ναυτικών, χρονική διάρκεια τηλεφωνημάτων, ημέρες διακοπών, ώρες εργασίας). Ένα δείγμα σπουδαστών μπορεί να εξετασθεί ως προς τις μεταβλητές: βάρος, ύψος, επίδοση, φύλλο, τόπος καταγωγής, εθνικότητα, επάγγελμα γονέων, αριθμός απουσιών, πλήθος οφειλομένων μαθημάτων.

Ένα δείγμα τηλεθεατών μπορεί να εξετασθεί ως προς τις μεταβλητές: φύλλο, ηλικία, μορφωτικό επίπεδο, ποιές ή και πόσες ώρες της ημέρας παρακολουθούν, τηλεοπτικός δίαυλος (κανάλι) ή δίαυλοι που προτιμούν, είδος εκπομπών που επιλέγουν (αθλητικές, σήριαλ, ενημερωτικές, ψυχαγωγικές).

Ένα δείγμα προϊόντων μπορεί να εξετασθεί ως προς τις μεταβλητές: τιμή, ποιότητα, εμφάνιση συσκευασίας (χάρτινη, πλαστική, δερμάτινη), βάρος συσκευασίας (κιλό, πεντάκιλο, δεκάκιλο).

Περιγραφική στατιστική.

Τα στατιστικά δεδομένα χωρίζονται σε ποιοτικά (qualitative) και ποσοτικά (quantitative). **Στατιστικός πίνακας** ονομάζεται ο πίνακας στον οποίο τοποθετούνται οι τιμές της μεταβλητής Χ, έτσι ώστε να είναι εύκολα κατανοητές και να διευκολύνεται η εξαγωγή αξιόπιστων συμπερασμάτων. Οι στατιστικοί πίνακες διακρίνονται σε γενικούς και ειδικούς.

Γενικοί ονομάζονται οι στατιστικοί πίνακες που περιέχουν όλες τις πληροφορίες που προκύπτουν από τη στατιστική έρευνα. Αποτελούν πηγή στατιστικών πληροφοριών.

Ειδικοί ονομάζονται οι στατιστικοί πίνακες που είναι συνοπτικοί και σαφείς. Τα στοιχεία που περιέχουν λαμβάνονται από τους γενικούς πίνακες. Κάθε στατιστικός πίνακας πρέπει να περιέχει:

- Τίτλο, που είναι γραμμένος στο πάνω μέρος του και δηλώνει συνοπτικά και με σαφήνεια το περιεχόμενο του.
- Επικεφαλίδες γραμμών και στηλών, που δηλώνουν συνοπτικά τη φύση και τις μονάδες μετρήσεως των δεδομένων.
- Κύριο σώμα, που περιέχει τα στατιστικά δεδομένα διαχωρισμένα σε γραμμές και στήλες.

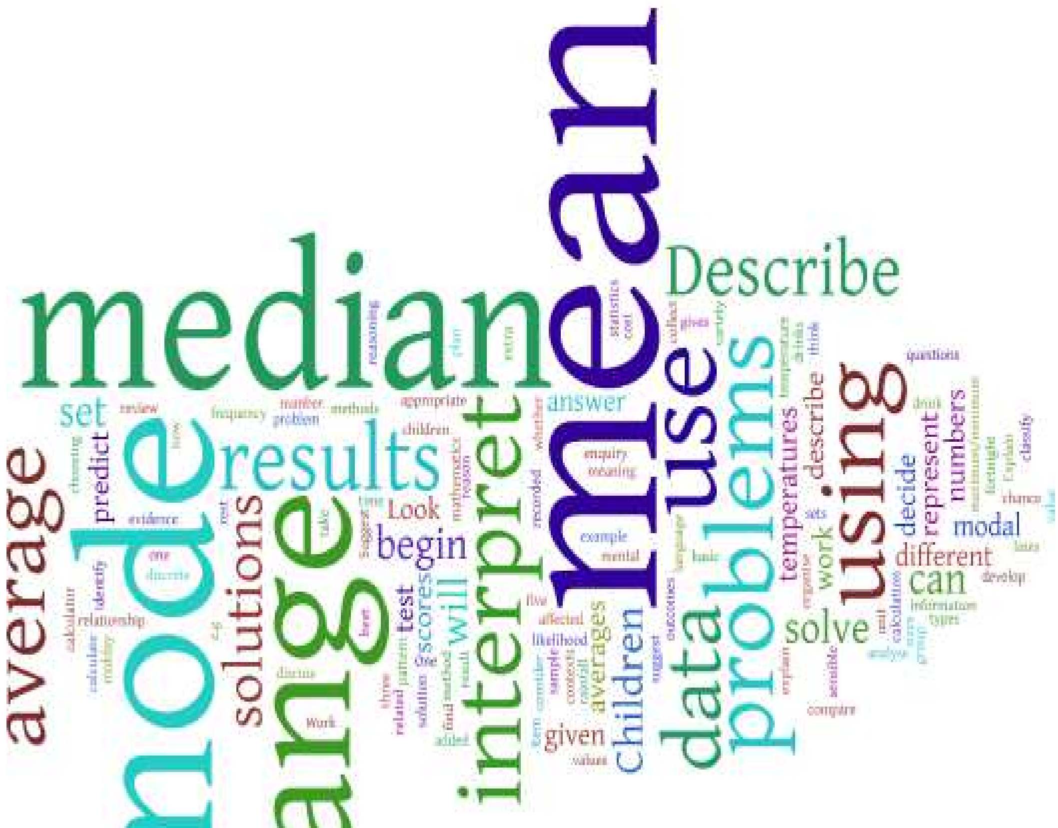
• Πηγή, γραμμένη στο κάτω μέρος του. Δείχνει την προέλευση των στατιστικών δεδομένων.

Ερωτήσεις σωστού – λάθους.

1. Ένα μικρό δείγμα είναι πάντα πιο αξιόπιστο από ένα μεγάλο δείγμα.
2. Ένα μεγάλο δείγμα είναι πάντα πιο αξιόπιστο από ένα μικρό δείγμα.
3. Το ύψος των ναυτικών ενός πλοίου είναι ποιοτική μεταβλητή.
4. Οι διακριτές μεταβλητές λαμβάνουν μόνο μεμονωμένες τιμές.
5. Το βάρος των υπαλλήλων μίας εταιρείας είναι ποσοτική συνεχής μεταβλητή.

Ερωτήσεις.

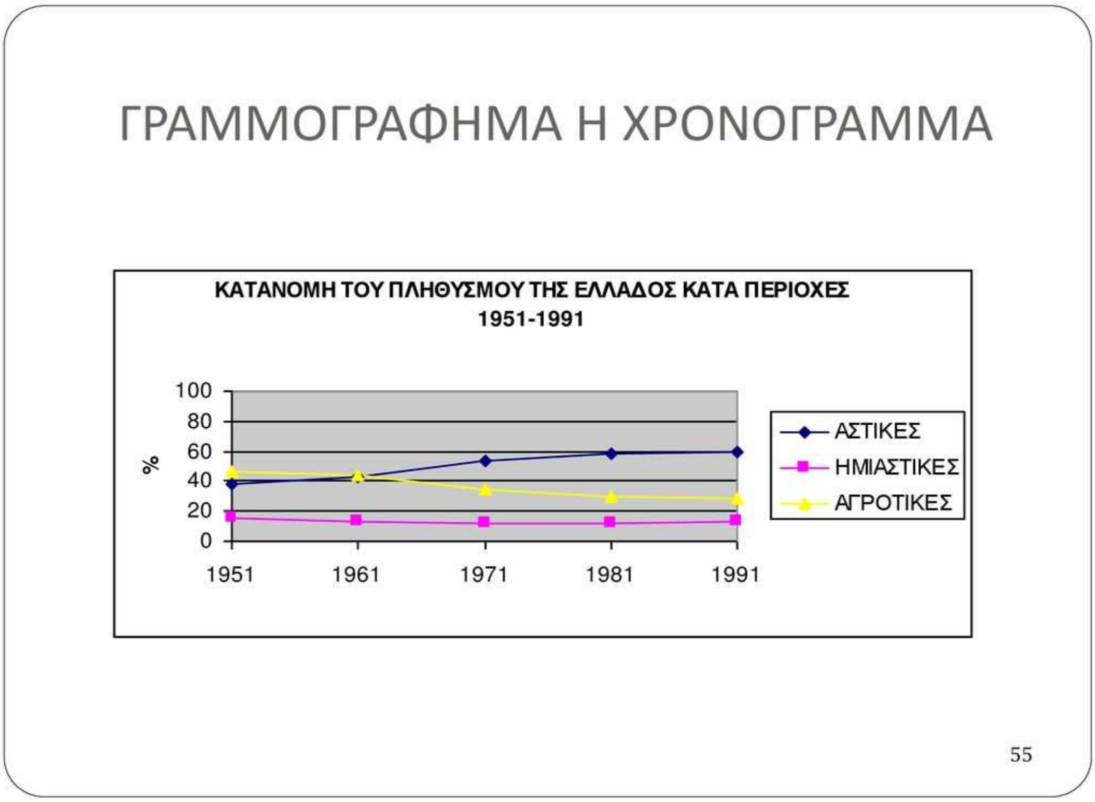
1. Με ποιον τρόπο θα επέλεγες ένα δείγμα από τους 200 σπουδαστές του Β΄ εξαμήνου, προκειμένου να εξάγεις συμπέρασμα για την επίδοση τους στα Μαθηματικά ΙΙ;

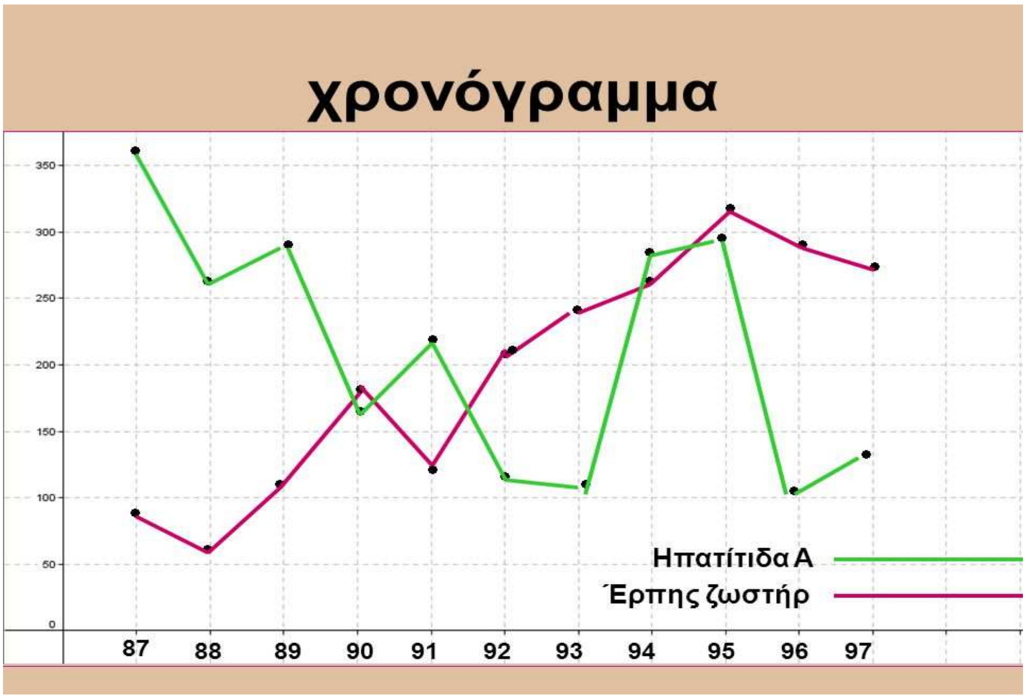


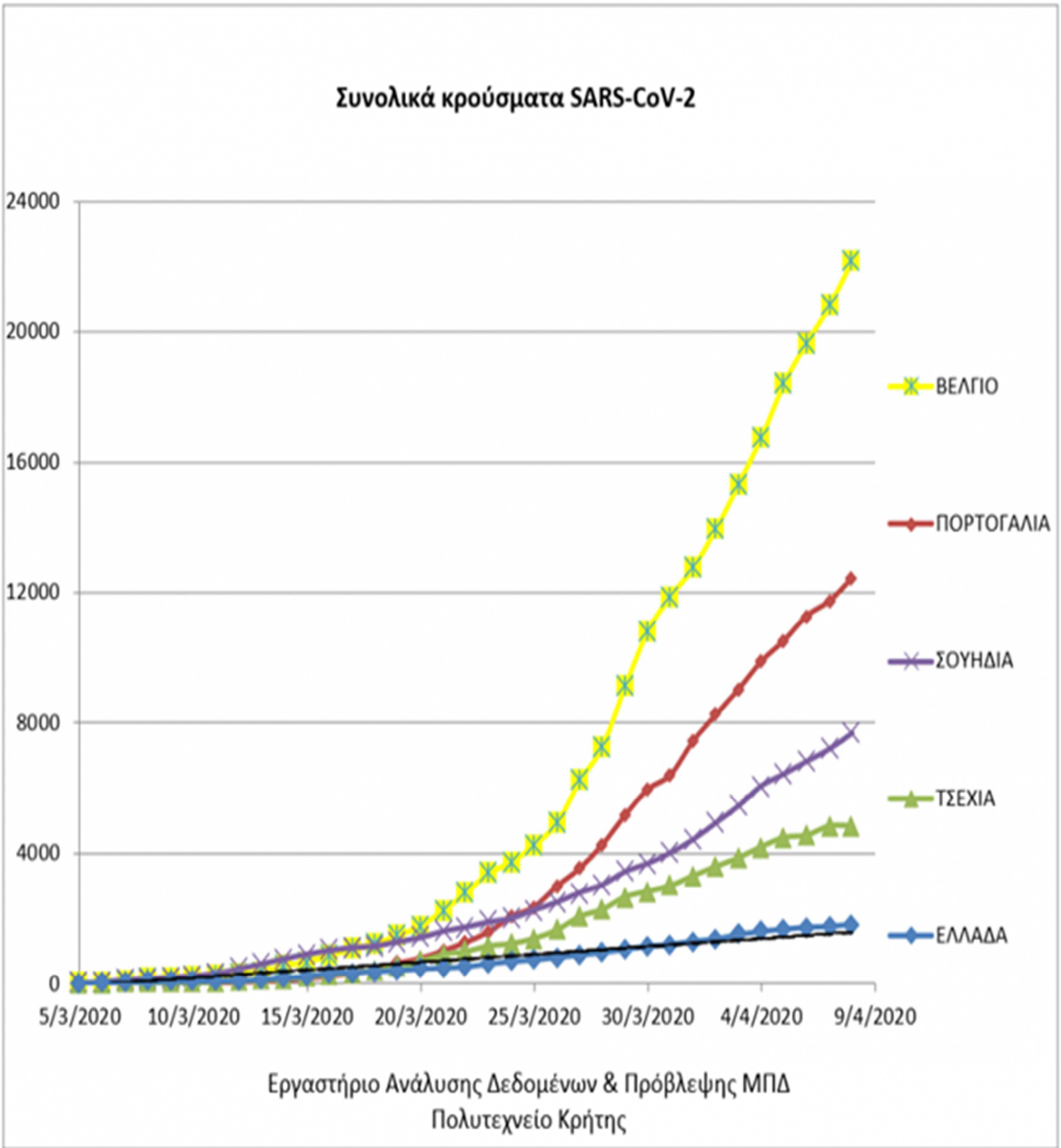
Γραμμογράφημα ή χρονόγραμμα.

Χρησιμοποιείται για να παραστήσομε γραφικά ένα φαινόμενο που εξελίσσεται με το πέρασμα του χρόνου (πυρετός ασθενούς, τιμή μετοχής, γεννήσεις, θάνατοι, κέρδη ή ζημίες επιχειρήσεως).

Στον οριζόντιο άξονα τοποθετούνται οι χρονικές στιγμές (ώρες, ημέρες, εβδομάδες, μήνες, έτη) και στον κάθετο άξονα οι τιμές της μεταβλητής που επιθυμούμε να μελετήσομε.






Συνολικά κρούσματα SARS-CoV-2
24000
20000
16000
12000
8000
4000
0
5/3/2020
10/3/2020
15/3/2020
20/3/2020
25/3/2020
30/3/2020
4/4/2020
9/4/2020
ΒΕΛΓΙΟ
ΠΟΡΤΟΓΑΛΙΑ
ΣΟΥΗΔΙΑ
ΤΣΕΧΙΑ
ΕΛΛΑΔΑ
Εργαστήριο Ανάλυσης Δεδομένων & Πρόβλεψης ΜΠΔ
Πολυτεχνείο Κρήτης

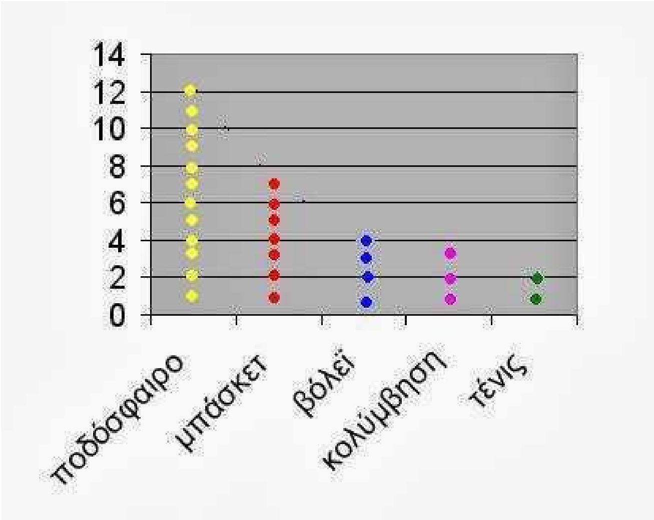

Εκατ. ευρώ
40
35
30
25
20
15
10
5
2002
2003
2004
2005
2006
2007
2008
Έξοδα
Έσοδα

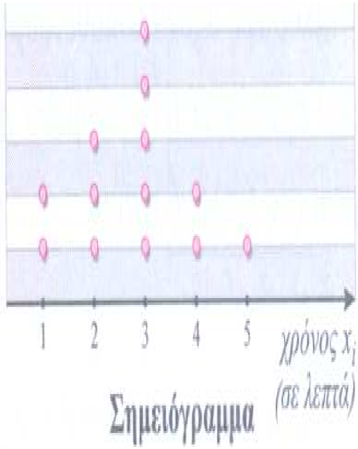

Αριθμός Κρουσμάτων
26 Φεβρουαρίου - 5 Απριλίου
20000
15000
10000
5000
0
26/02
29/02
03/03
06/03
09/03
12/03
15/03
18/03
21/03
26/03
29/03
30/03
3/04
5/04
19.691
Βέλγιο
Πληθυσμός: 11,04 εκατ.
17.851
Ολλανδία
Πληθυσμός: 17,18 εκατ.
11.278
Πορτογαλία
Πληθυσμός: 10,3 εκατ.
1.735
Ελλάδα
Πληθυσμός: 10,8 εκατ.

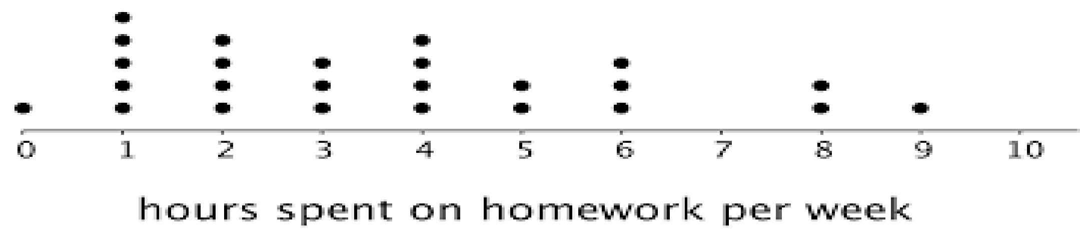

(γ)
Σύγκριση συνολικών κρουσμάτων
Αριθμός κρουσμάτων
6,000
5,000
4,000
3,000
2,000
1,000
0
Ισπανία
Ιταλία
Ελβετία
Γαλλία
Γερμανία
Ελλάδα
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36
Ημέρες από τα πρώτα κρούσματα

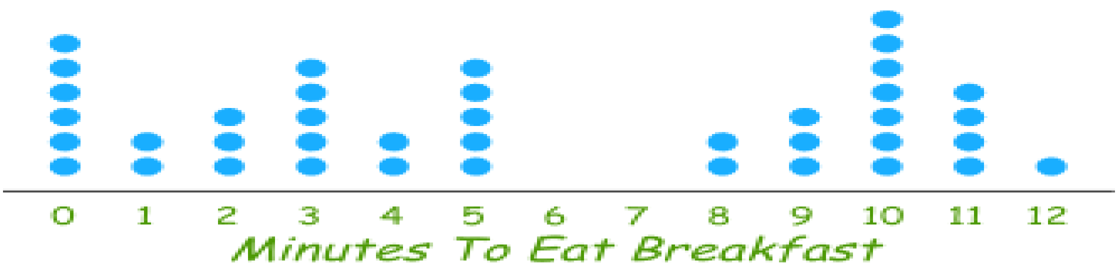
Σημειόγραμμα (Dot diagram).

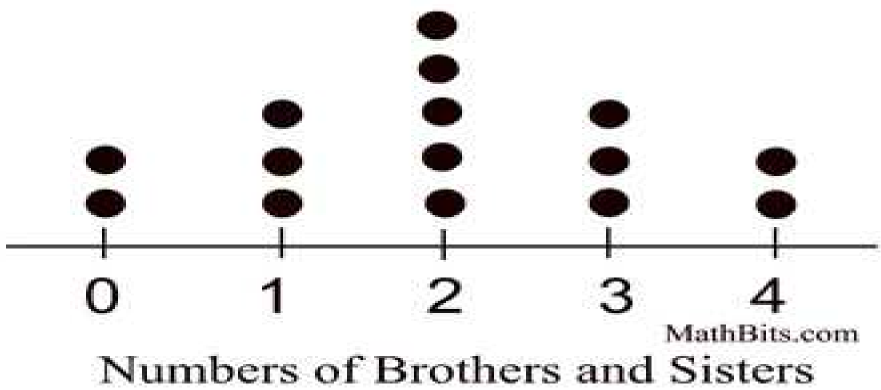
Χρησιμοποιείται όταν το μέγεθος του δείγματος είναι μικρό (λίγες παρατηρήσεις). Στον οριζόντιο άξονα τοποθετούμε τις τιμές της μεταβλητής και πάνω από κάθε τιμή βάζομε κατακόρυφα τόσες τελείες όση και η αντίστοιχη συχνότητα.











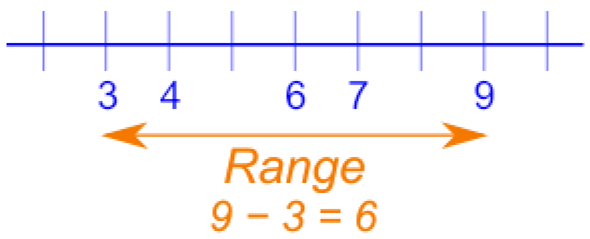
Εύρος ή κύμανση.

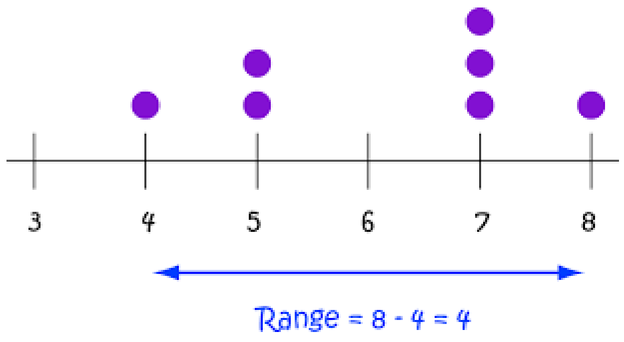
Εύρος ή κύμανση (Range) R ονομάζεται η διαφορά της ελάχιστης από τη μέγιστη παρατήρηση. Δηλαδή $R=\beta-\alpha$, όπου α η μικρότερη και β η μεγαλύτερη παρατήρηση.

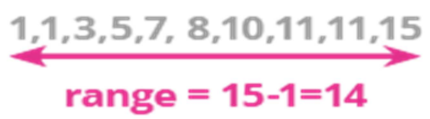
Για ομαδοποιημένα δεδομένα, το εύρος δίνεται από τη διαφορά του κατωτέρου ορίου της πρώτης κλάσεως, από το ανώτερο όριο της τελευταίας κλάσεως.

Δεν είναι ικανοποιητικό μέτρο διασποράς, διότι εξαρτάται μόνο από τις ακραίες τιμές των παρατηρήσεων. Έτσι, αν οι δυο ακραίες τιμές είναι αρκετά απομακρυσμένες από τις άλλες, δίνουν ψεύτικη εικόνα.

Πλεονεκτήματα του εύρους.	**Μειονεκτήματα του εύρους.**
Χρησιμοποιείται αρκετά στον έλεγχο ποιότητας.	Δε θεωρείται αξιόπιστο μέτρο διασποράς.
Είναι πολύ απλό στον υπολογισμό.	Δε χρησιμοποιείται για περαιτέρω στατιστική ανάλυση.
Μπορεί να χρησιμοποιηθεί για την εκτίμηση της τυπικής απόκλισης.	Δεν υπολογίζεται για ποιοτικά δεδομένα.







How to Calculate the Range:

Example Values

12, 7, 17, 9, 16, 29, 8, 4, 15, 7, 29, 6, 17, 2, 5, 12, 8, 17, 6, 6, 17, 17

Step 1:

Sort all the values from low to high.

2, 4, 5, 6, 6, 6, 7, 7, 8, 9, 12, 12, 15, 16, 17, 17, 17, 17, 17, 29

Step 2:

Subtract the lowest value from the highest value to determine the range.

2, 4, 5, 6, 6, 6, 7, 7, 8, 9, 12, 12, 15, 16, 17, 17, 17, 17, 17, 29

Range = 29 - 2 = **27**

Εφαρμογή 1. Συμπληρώστε τον παρακάτω πίνακα που αφορά την εξέταση δείγματος (20) είκοσι πωλητών ιδιωτικής χρήσεως αυτοκινήτων αντιπροσωπείας ως προς τις πωλήσεις που σημείωσαν σε διάστημα μίας εβδομάδος.

Πλήθος πωληθέντων αυτοκινήτων	Συχνότητα ν_i	Αθροιστική συχνότητα N_i	Σχετική συχνότητα f_i	Αθροιστική σχετική συχνότητα F_i	Σχετική επί τοις % συχνότητα $f_i\%$	Επί τοις % αθροιστική σχετική συχνότητα $F_i\%$
$x_1 = 0$	$\nu_1 = 10$	$N_1 =$	$f_1 =$	$F_1 =$	$f_1\% =$	$F_1\% =$
$x_2 = 1$	$\nu_2 = 5$	$N_2 =$	$f_2 =$	$F_2 =$	$f_2\% =$	$F_2\% =$
$x_3 = 2$	$\nu_3 = 2$	$N_3 =$	$f_3 =$	$F_3 =$	$f_3\% =$	$F_3\% =$
$x_4 = 3$	$\nu_4 = 2$	$N_4 =$	$f_4 =$	$F_4 =$	$f_4\% =$	$F_4\%$
$x_5 = 4$	$\nu_5 = 1$	$N_5 =$	$f_5 =$	$F_5 =$	$f_5\% =$	$F_5\% =$
Σύνολο	$\nu = \sum_{i=1}^{5} \nu_i = 20$		$\sum_{i=1}^{5} f_i = 1$		$\sum_{i=1}^{5} f_i\% = 100\%$	

Λύση.

Πλήθος πωληθέντων αυτοκινήτων.	Συχνότητα ν_i	Αθροιστική συχνότητα N_i	Σχετική συχνότητα f_i	Αθροιστική σχετική συχνότητα F_i	Επί τοις % σχετική συχνότητα $f_i\%$	Αθροιστική επί τοις % σχετική συχνότητα $F_i\%$
$x_1 = 0$	$\nu_1 = 10$	$N_1 = 10$	$f_1 = 0,5$	$F_1 = 0,50$	$f_1\% = 50\%$	$F_1\% = 50\%$
$x_2 = 1$	$\nu_2 = 5$	$N_2 = 15$	$f_2 = 0,25$	$F_2 = 0,75$	$f_2\% = 25\%$	$F_2\% = 75\%$
$x_3 = 2$	$\nu_3 = 2$	$N_3 = 17$	$f_3 = 0,10$	$F_3 = 0,85$	$f_3\% = 10\%$	$F_3\% = 85\%$
$x_4 = 3$	$\nu_4 = 2$	$N_4 = 19$	$f_4 = 0,10$	$F_4 = 0,95$	$f_4\% = 10\%$	$F_4\% = 95\%$
$x_5 = 4$	$\nu_5 = 1$	$N_5 = 20$	$f_5 = 0,05$	$F_5 = 1$	$f_5\% = 5\%$	$F_5\% = 100\%$
Σύνολο	$\nu = \sum_{i=1}^{5} \nu_i = 20$		$\sum_{i=1}^{5} f_i = 1$		$\sum_{i=1}^{5} f_i\% = 100\%$	

$N_1 = v_1 = 10$

$N_2 = v_1 + v_2 = N_1 + v_2 = 10 + 5 = 15$

$N_3 = v_1 + v_2 + v_3 = N_2 + v_3 = 15 + 2 = 17$

$N_4 = v_1 + v_2 + v_3 + v_4 = N_3 + v_4 = 17 + 2 = 19$

$N_5 = N_4 + v_5 = 19 + 1 = 20 = v$

$f_1 = \dfrac{v_1}{v} = \dfrac{10}{20} = 0,5$

$f_2 = \dfrac{v_2}{v} = \dfrac{5}{20} = \dfrac{1}{4} = 0,25$

$f_3 = \dfrac{v_3}{v} = \dfrac{2}{20} = \dfrac{1}{10} = 0,1$

$f_4 = \dfrac{v_4}{v} = \dfrac{2}{20} = \dfrac{1}{10} = 0,1$

$f_5 = \dfrac{v_5}{v} = \dfrac{1}{20} = \dfrac{0,5}{10} = 0,05$

$F_1 = f_1 = 0,50$

$F_2 = f_1 + f_2 = F_1 + f_2 = 0,50 + 0,25 = 0,75$

$F_3 = \sum_{i=1}^{3} f_i = F_2 + f_3 = 0,75 + 0,10 = 0,85$

$F_4 = \sum_{i=1}^{4} f_i = F_3 + f_4 = 0,85 + 0,10 = 0,95$

$F_5 = \sum_{i=1}^{5} f_i = F_4 + f_5 = 0,95 + 0,05 = 1$

$f_1\% = \dfrac{v_1}{v}100 = \dfrac{10}{20}100 = 50$

$f_2\% = \dfrac{v_2}{v}100 = \dfrac{5}{20}100 = \dfrac{1}{4}100 = 25$

$f_3\% = \dfrac{v_3}{v}100 = \dfrac{2}{20}100 = \dfrac{1}{10}100 = 10$

$f_4\% = \dfrac{v_4}{v}100 = \dfrac{2}{20}100 = \dfrac{1}{10}100 = 10$

$f_5\% = \dfrac{v_5}{v}100 = \dfrac{1}{20}100 = \dfrac{0,5}{10}100 = 5$

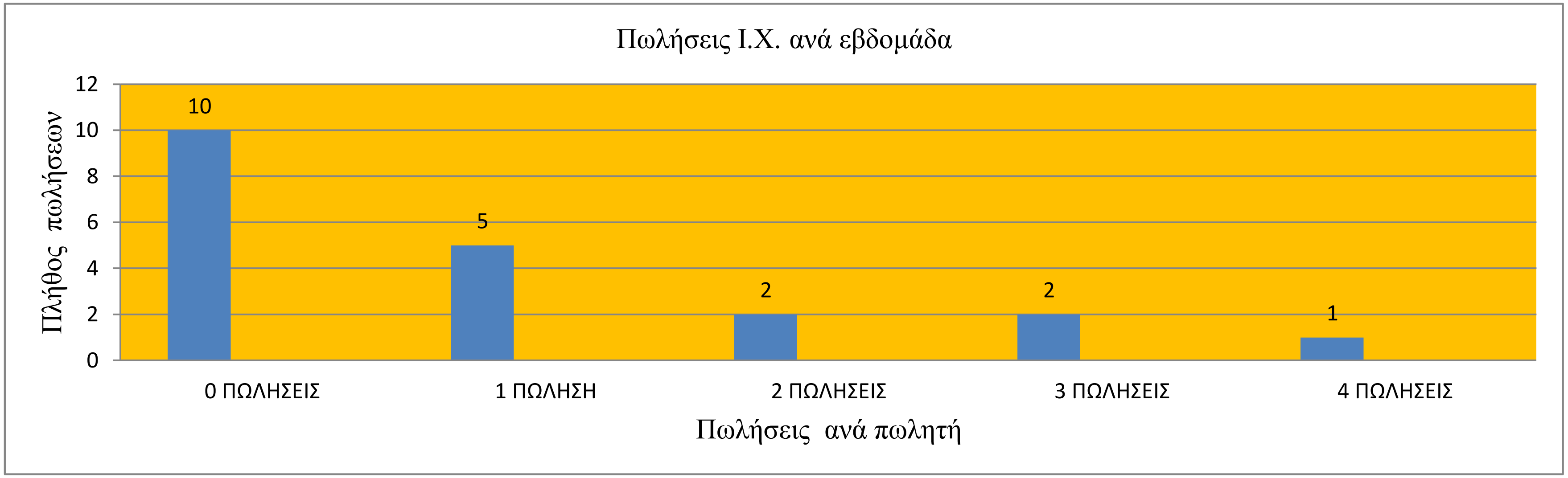
$F_1\% = f_1\% = 50\%$

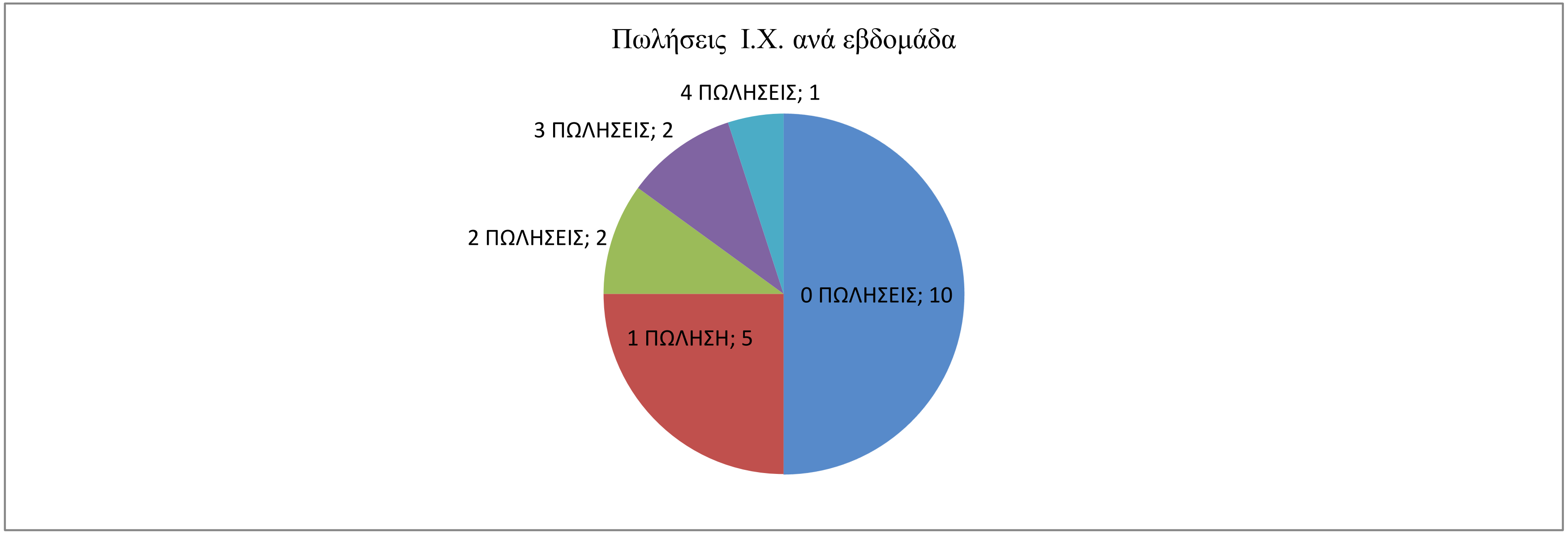
$F_2\% = f_1\% + f_2\% = F_1\% + f_2\% = 50\% + 25\% = 75\%$

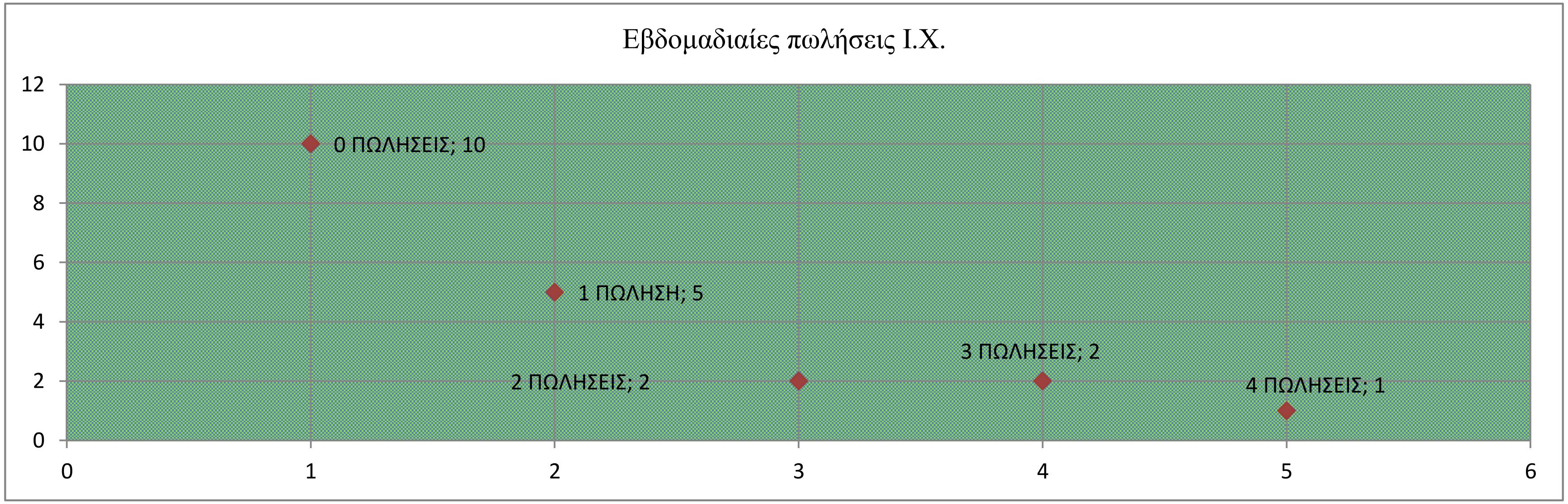
$F_3\% = \sum_{i=1}^{3} f_i\% = F_2\% + f_3\% = 75\% + 10\% = 85\%$

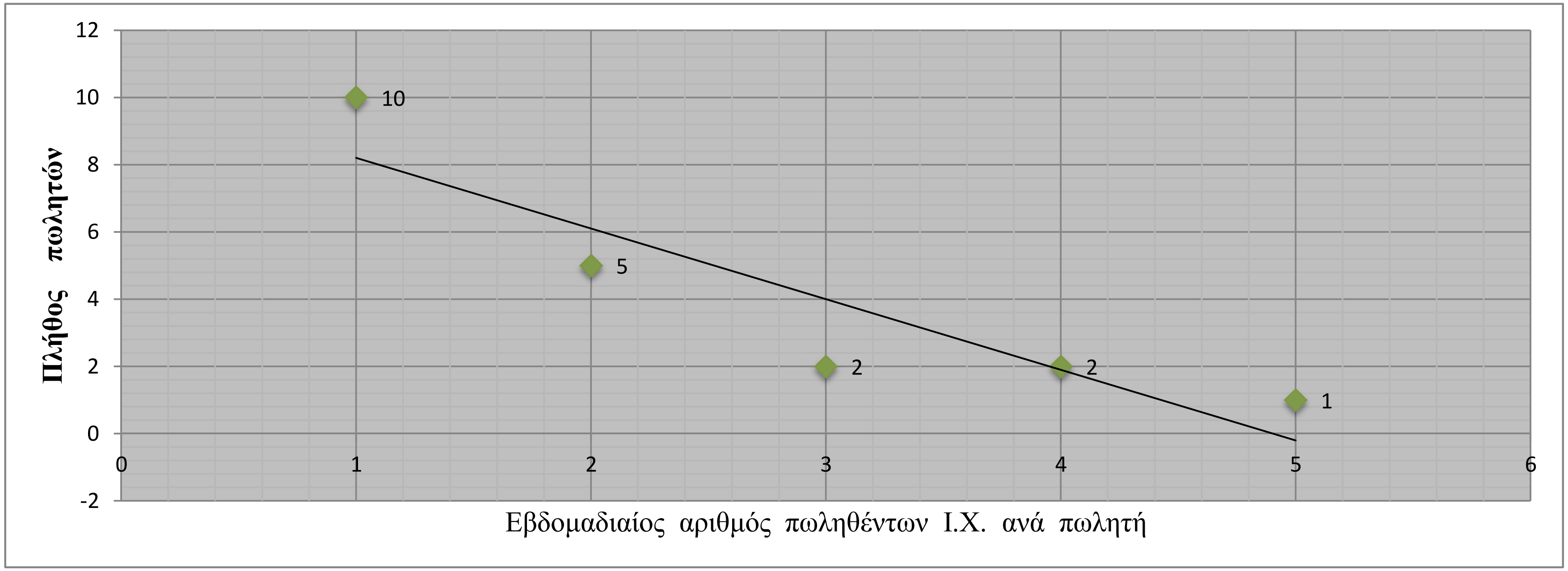
$F_4\% = \sum_{i=1}^{4} f_i\% = F_3\% + f_4\% = 85\% + 10\% = 95\%$

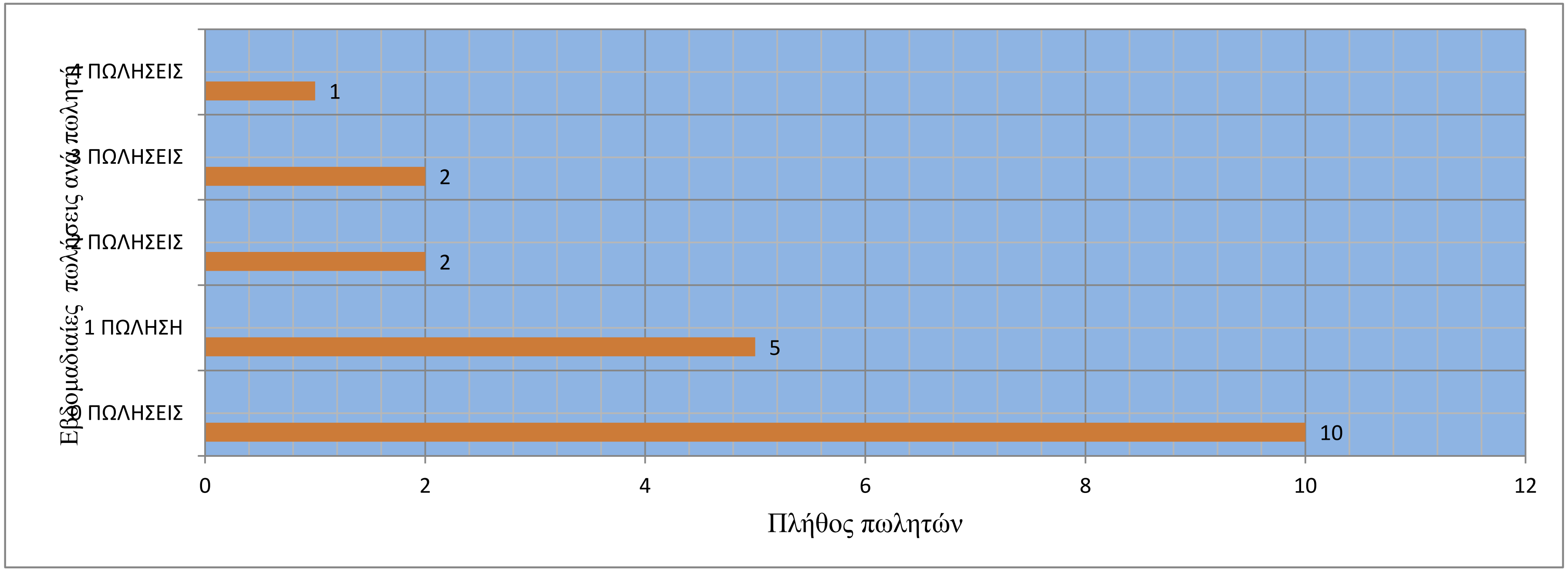
$F_5\% = \sum_{i=1}^{5} f_i\% = F_4\% + f_5\% = 95\% + 5\% = 100\%$


Πωλήσεις Ι.Χ. ανά εβδομάδα
Πλήθος πωλήσεων
12
10
8
6
4
2
0
10
5
2
2
1
0 ΠΩΛΗΣΕΙΣ
1 ΠΩΛΗΣΗ
2 ΠΩΛΗΣΕΙΣ
3 ΠΩΛΗΣΕΙΣ
4 ΠΩΛΗΣΕΙΣ
Πωλήσεις ανά πωλητή


Πωλήσεις Ι.Χ. ανά εβδομάδα
4 ΠΩΛΗΣΕΙΣ; 1
3 ΠΩΛΗΣΕΙΣ; 2
2 ΠΩΛΗΣΕΙΣ; 2
0 ΠΩΛΗΣΕΙΣ; 10
1 ΠΩΛΗΣΗ; 5


Εβδομαδιαίες πωλήσεις Ι.Χ.
0 ΠΩΛΗΣΕΙΣ; 10
1 ΠΩΛΗΣΗ; 5
2 ΠΩΛΗΣΕΙΣ; 2
3 ΠΩΛΗΣΕΙΣ; 2
4 ΠΩΛΗΣΕΙΣ; 1


Πλήθος πωλητών
10
5
2
2
1
Εβδομαδιαίος αριθμός πωληθέντων Ι.Χ. ανά πωλητή



Εβδομαδιαίες πωλήσεις ανά πωλητή
4 ΠΩΛΗΣΕΙΣ
3 ΠΩΛΗΣΕΙΣ
2 ΠΩΛΗΣΕΙΣ
1 ΠΩΛΗΣΗ
0 ΠΩΛΗΣΕΙΣ
1
2
2
5
10
0
2
4
6
8
10
12
Πλήθος πωλητών

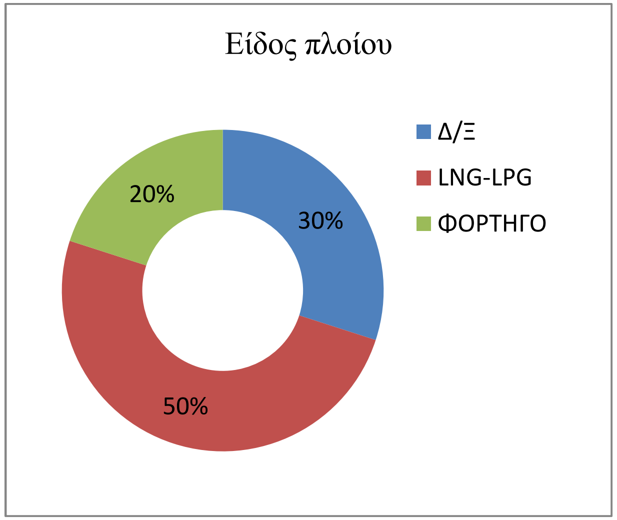
Εφαρμογή 2. ν το πλήθος σπουδαστές ΑΕΝ ρωτήθηκαν για το είδος πλοίου που επιθυμούν να ναυτολογηθούν και τα αποτελέσματα φαίνονται στον παρακάτω πίνακα. Να συμπληρωθεί ο πίνακας.

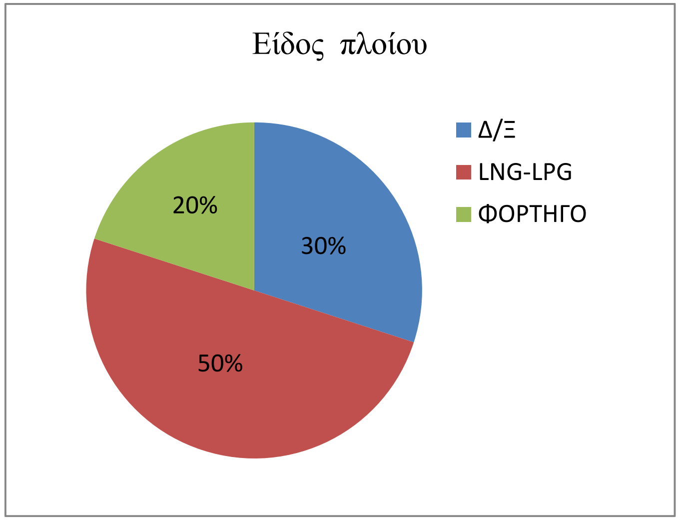
x_i	ν_i	N_i	$f_i\%$	$F_i\%$	f_i	F_i
$x_1 =$ Tanker	$\nu_1 =$	$N_1 =$	$f_1\% =$	$F_1\% =$	$f_1 =$	$F_1 =$
$x_2 =$ LNG-LPG	$\nu_2 = 75$	$N_2 =$	$f_2\% =$	$F_2\% = 80\%$	$f_2 =$	$F_2 =$
$x_3 =$ Φορτηγό	$\nu_3 = 30$	$N_3 =$	$f_3\% =$	$F_3\% =$	$f_3 =$	$F_3 =$
Σύνολο	$\nu = \sum_{i=1}^{3} \nu_i =$		$\sum_{i=1}^{3} f_i\% = 100\%$		$\sum_{i=1}^{3} f_i = 1$	

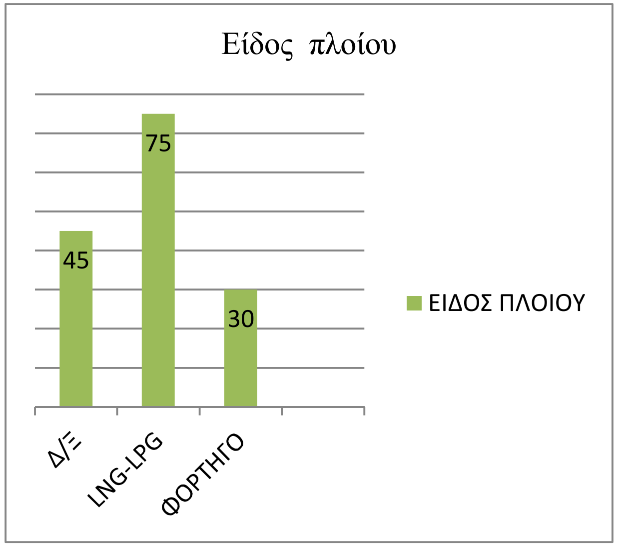
Λύση.

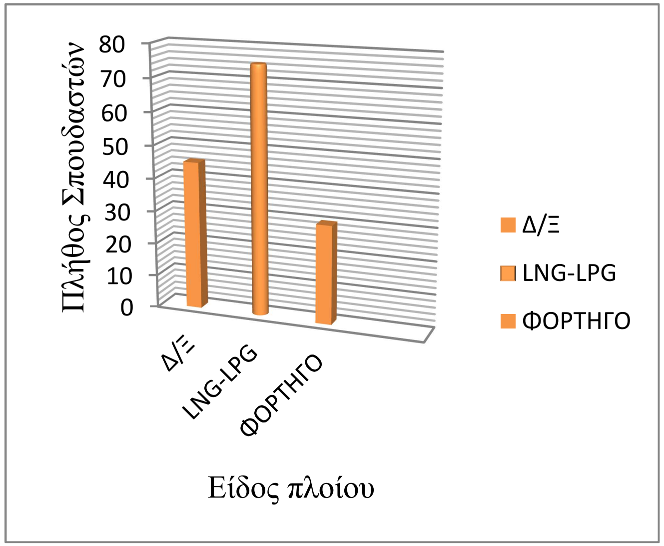
Είναι $\frac{\nu_1+75}{\nu} = \frac{80}{100} \Leftrightarrow \frac{\nu_1+75}{\nu} = \frac{8}{10} \Leftrightarrow \frac{\nu_1+75}{\nu} = \frac{4}{5} \Leftrightarrow \frac{\nu_1+75}{\nu_1+105} = \frac{4}{5} \Leftrightarrow 5(\nu_1+75) = 4(\nu_1+105) \Leftrightarrow 5\nu_1 + 375 = 4\nu_1 + 420 \Leftrightarrow \nu_1 = 45$

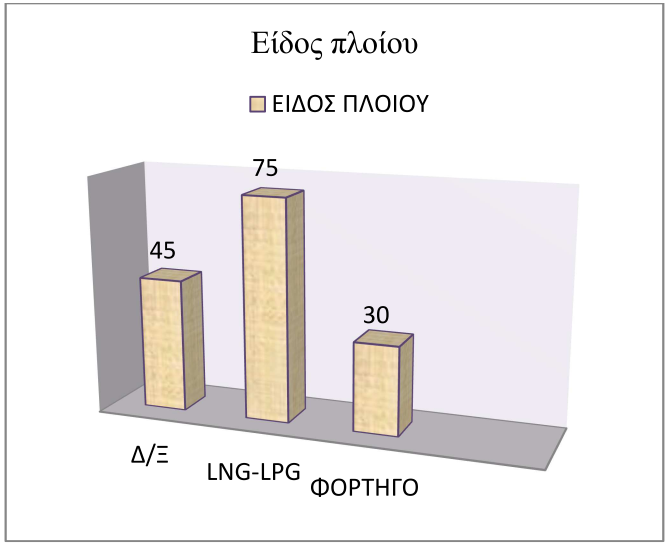
x_i	ν_i	N_i	$f_i\%$	$F_i\%$	f_i	F_i
$x_1 =$ Tanker	$\nu_1 = 45$	$N_1 = 45$	$f_1\% = 30\%$	$F_1\% = 30\%$	$f_1 = 0,3$	$F_1 = 0,3$
$x_2 =$ LNG-LPG	$\nu_2 = 75$	$N_2 = 120$	$f_2\% = 50\%$	$F_2\% = 80\%$	$f_2 = 0,5$	$F_2 = 0,8$
$x_3 =$ Φορτηγό	$\nu_3 = 30$	$N_3 = 150$	$f_3\% = 20\%$	$F_3\% = 100\%$	$f_3 = 0,2$	$F_3 = 1$
Σύνολο	$\nu = \sum_{i=1}^{3} \nu_i = 150$		$\sum_{i=1}^{3} f_i\% = 100\%$		$\sum_{i=1}^{3} f_i = 1$	

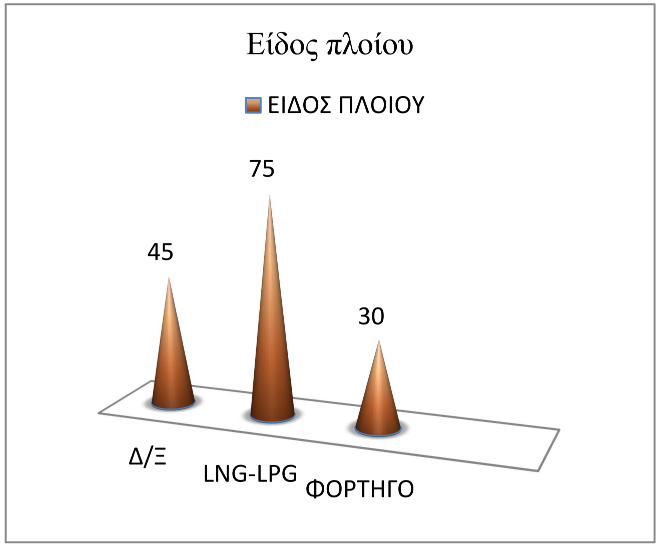

Είδος πλοίου
Δ/Ξ
LNG-LPG
ΦΟΡΤΗΓΟ
30%
50%
20%

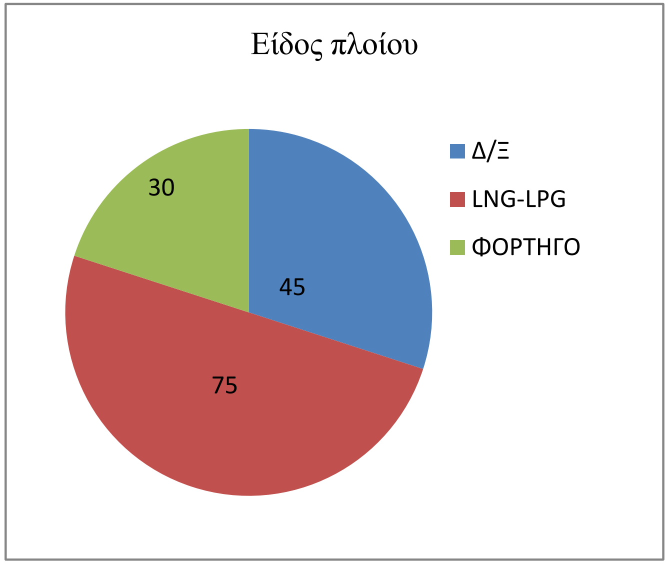

Είδος πλοίου
Δ/Ξ
LNG-LPG
ΦΟΡΤΗΓΟ
30%
50%
20%

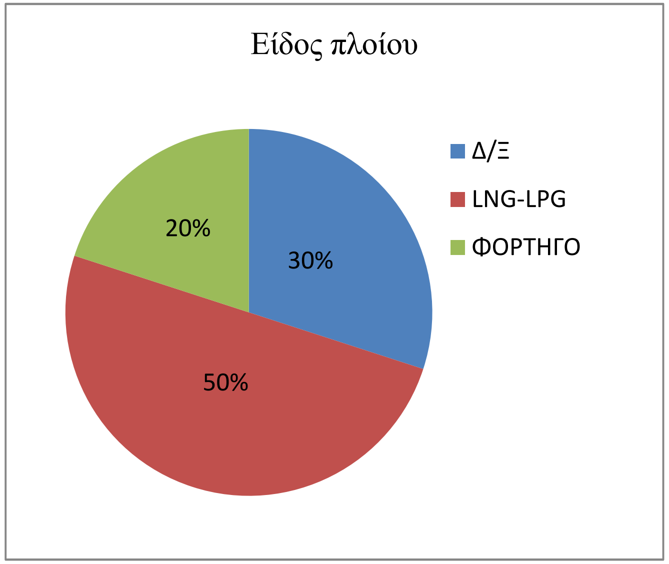

Είδος πλοίου
45
75
30
Δ/Ξ
LNG-LPG
ΦΟΡΤΗΓΟ
ΕΙΔΟΣ ΠΛΟΙΟΥ

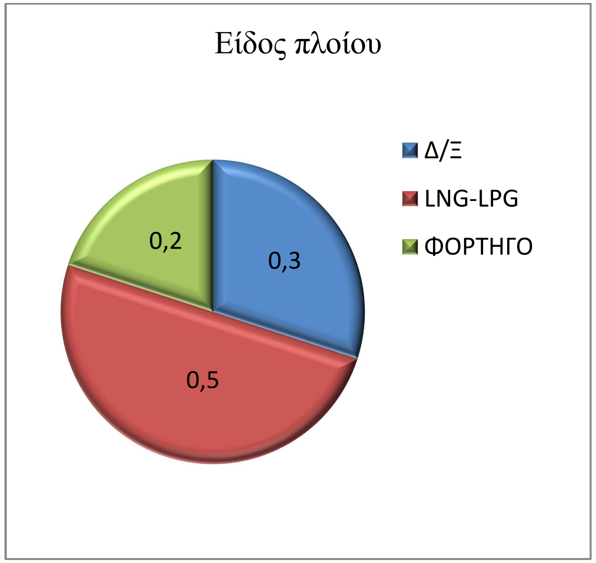

Πλήθος Σπουδαστών
0
10
20
30
40
50
60
70
80
Δ/Ξ
LNG-LPG
ΦΟΡΤΗΓΟ
Είδος πλοίου


Είδος πλοίου
ΕΙΔΟΣ ΠΛΟΙΟΥ
45
75
30
Δ/Ξ
LNG-LPG
ΦΟΡΤΗΓΟ


Είδος πλοίου
ΕΙΔΟΣ ΠΛΟΙΟΥ
45
75
30
Δ/Ξ
LNG-LPG
ΦΟΡΤΗΓΟ


Είδος πλοίου
30
45
75
Δ/Ξ
LNG-LPG
ΦΟΡΤΗΓΟ


Είδος πλοίου
20%
30%
50%
Δ/Ξ
LNG-LPG
ΦΟΡΤΗΓΟ


Είδος πλοίου
Δ/Ξ
LNG-LPG
ΦΟΡΤΗΓΟ
0,2
0,3
0,5

Εφαρμογή 3. Να συμπληρωθεί ο πίνακας προκύπτει όταν είκοσι (20) σπουδαστές ερωτώνται για το πλήθος των ωρών που μελετούν ημερησίως.

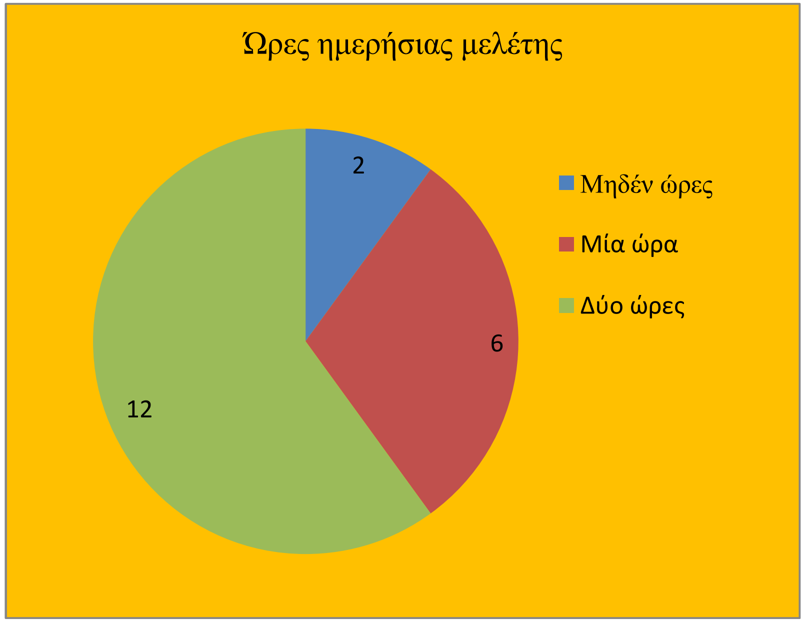
x_i	ν_i	N_i	f_i	F_i	$f_i\%$	$F_i\%$	$\widehat{a}_i$
$x_1=0$	$\nu_1=2$	$N_1=$	$f_1=$	$F_1=$	$f_1\%=$	$F_1\%=$	$\widehat{a}_1=$
$x_2=1$	$\nu_2=$	$N_2=$	$f_2=0,3$	$F_2=$	$f_2\%=$	$F_2\%=$	$\widehat{a}_2=$
$x_3=2$	$\nu_3=$	$N_3=$	$f_3=$	$F_3=$	$f_3\%=$	$F_3\%=$	$\widehat{a}_3=$
Σύνολο	$\nu=\sum_{i=1}^{3}\nu_i=20$		$\sum_{i=1}^{3}f_i=$		$\sum_{i=1}^{3}f_i\%=$		$\sum_{i=1}^{3}\widehat{a}_i=$

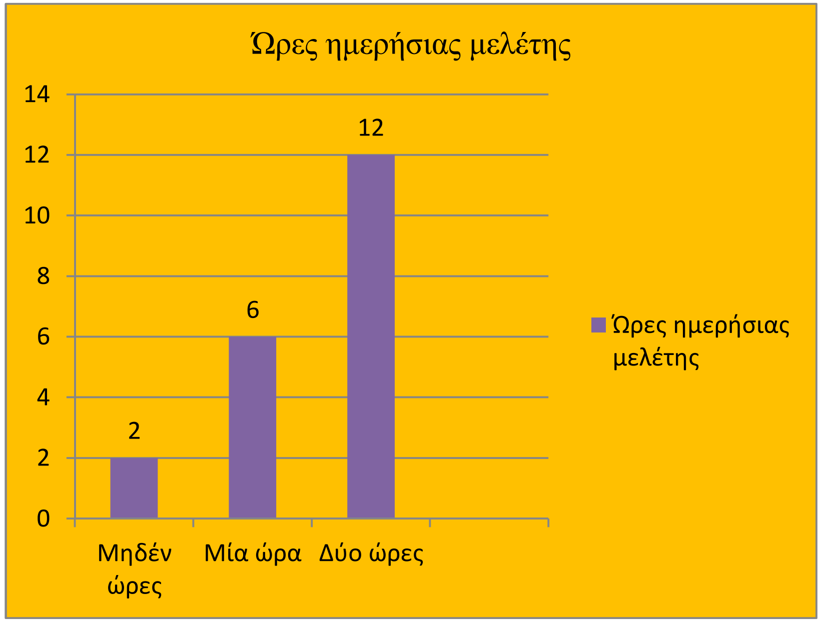
Λύση.

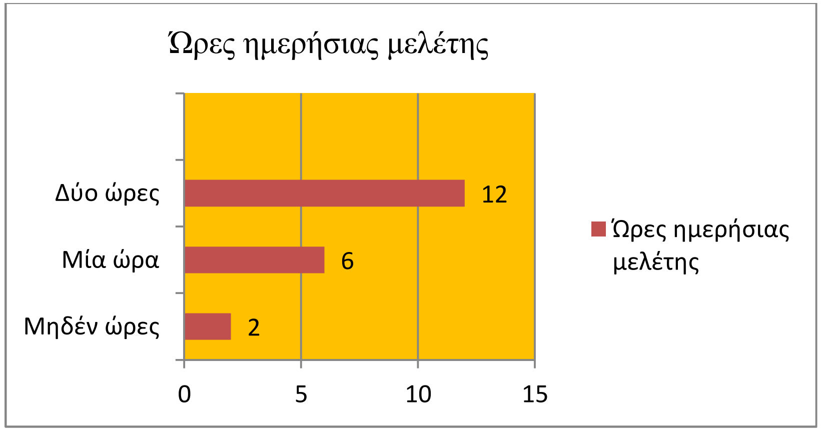
$$f_1=\frac{\nu_1}{\nu}=\frac{2}{20}=\frac{1}{10}=0,1 \qquad f_3=1-f_1-f_2=1-0,1-0,3=0,6 \qquad \widehat{a}_1=\frac{\nu_1}{\nu}360^0=\frac{2}{20}360^0=\frac{1}{10}360^0=36^0$$

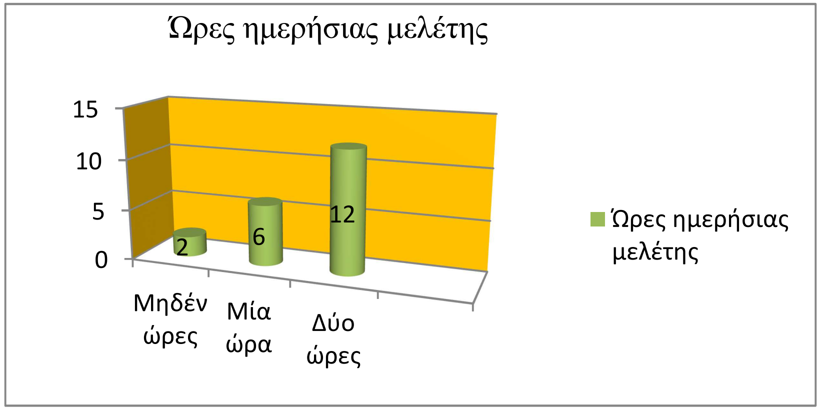
$$\widehat{a}_2=\frac{\nu_2}{\nu}360^0=\frac{6}{20}360^0=\frac{3}{10}360^0=3\cdot 36^0=108^0 \qquad \widehat{a}_3=\frac{\nu_3}{\nu}360^0=\frac{12}{20}360^0=\frac{6}{10}360^0=6\cdot 36^0=216^0$$

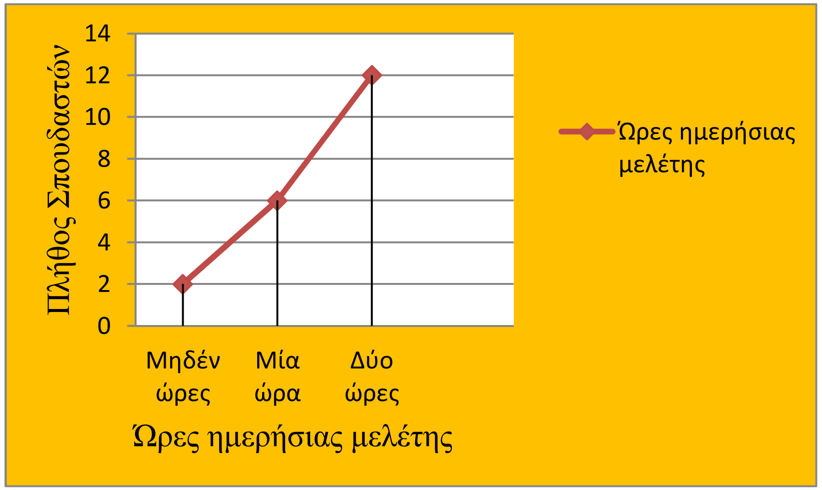
x_i	ν_i	N_i	f_i	F_i	$f_i\%$	$F_i\%$	$\widehat{a}_i$
$x_1=0$	$\nu_1=2$	$N_1=2$	$f_1=0,1$	$F_1=0,1$	$f_1\%=10\%$	$F_1\%=10\%$	$\widehat{a}_1=36^0$
$x_2=1$	$\nu_2=6$	$N_2=8$	$f_2=0,3$	$F_2=0,4$	$f_2\%=30\%$	$F_2\%=40\%$	$\widehat{a}_2=108^0$
$x_3=2$	$\nu_3=12$	$N_3=20$	$f_3=0,6$	$F_3=1$	$f_3\%=60\%$	$F_3\%=100\%$	$\widehat{a}_3=216^0$
Σύνολο	$\nu=\sum_{i=1}^{3}\nu_i=20$		$\sum_{i=1}^{3}f_i=1$		$\sum_{i=1}^{3}f_i\%=100\%$		$\sum_{i=1}^{3}\widehat{a}_i=360^0$


Ώρες ημερήσιας μελέτης
2
6
12
Μηδέν ώρες
Μία ώρα
Δύο ώρες


Ώρες ημερήσιας μελέτης
14
12
10
8
6
4
2
0
2
6
12
Μηδέν ώρες
Μία ώρα
Δύο ώρες
Ώρες ημερήσιας μελέτης


Ώρες ημερήσιας μελέτης
Δύο ώρες
Μία ώρα
Μηδέν ώρες
12
6
2
0
5
10
15
Ώρες ημερήσιας μελέτης


Ώρες ημερήσιας μελέτης
15
10
5
0
2
6
12
Μηδέν ώρες
Μία ώρα
Δύο ώρες
Ώρες ημερήσιας μελέτης


Πλήθος Σπουδαστών
14
12
10
8
6
4
2
0
Μηδέν ώρες
Μία ώρα
Δύο ώρες
Ώρες ημερήσιας μελέτης
Ώρες ημερήσιας μελέτης

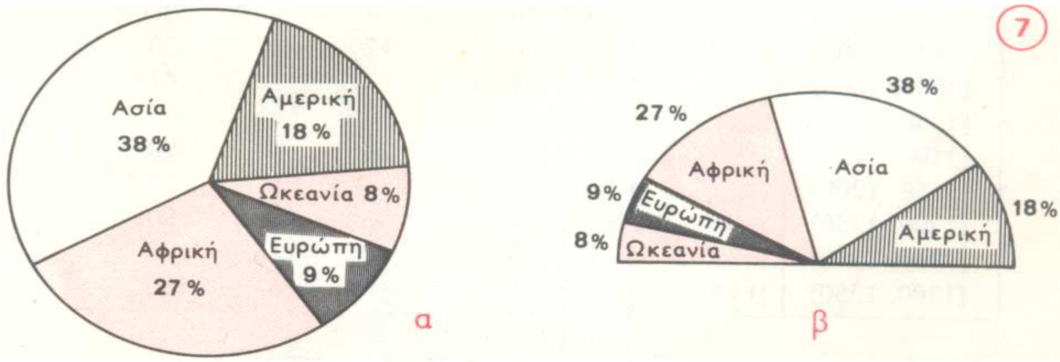

Ώρες ημερήσιας μελέτης
Πλήθος Σπουδαστών
14
12
10
8
6
4
2
0
Μηδέν ώρες
Μία ώρα
Δύο ώρες
Ώρες ημερήσιας μελέτης
2
6
12

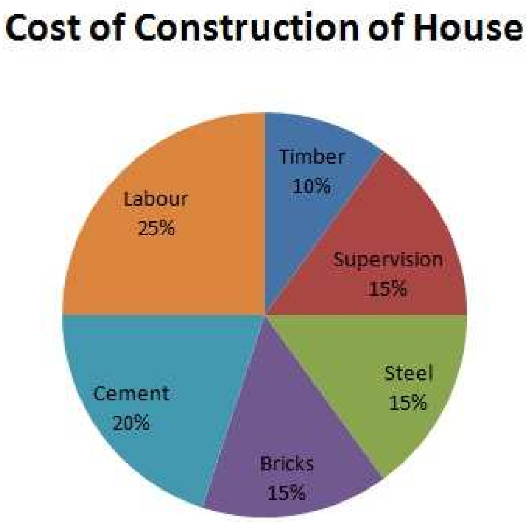

Ώρες ημερήσιας μελέτης
2
6
12
Μηδέν ώρες
Μία ώρα
Δύο ώρες

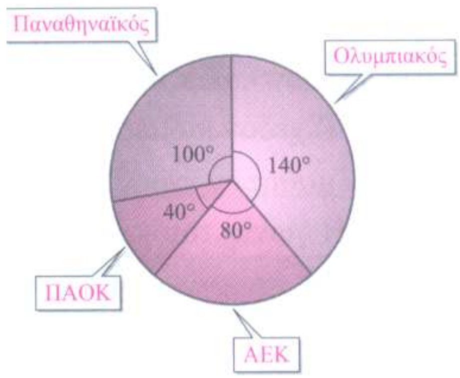
Κυκλικό διάγραμμα (piechart).

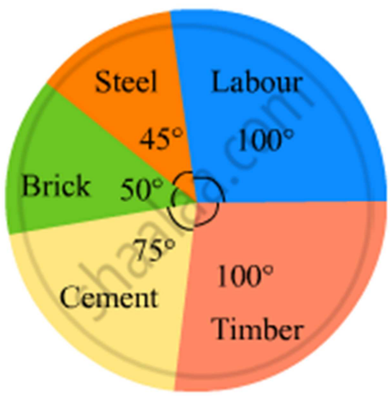
Χρησιμοποιείται για τη γραφική παράσταση ποσοτικών και κυρίως ποιοτικών δεδομένων. Αποτελείται από έναν κυκλικό δίσκο, χωρισμένο σε κυκλικούς τομείς, των οποίων τα εμβαδά ή ισοδύναμα τα τόξα είναι ανάλογα προς τις αντίστοιχες συχνότητες ν_i ή τις σχετικές συχνότητες f_i των τιμών x_i της μεταβλητής.

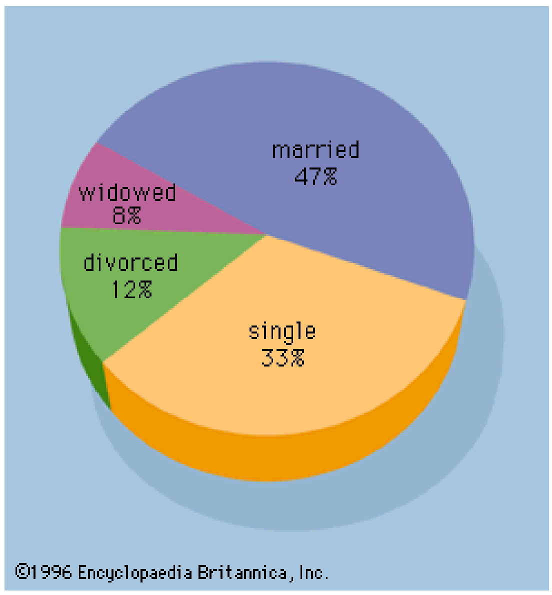
Αν a_i είναι το τόξο που αντιστοιχεί στην τιμή x_i, ισχύει ότι $a_i = \frac{\nu_i}{\nu} \cdot 360^0 = f_i \cdot 360^0$, αν πρόκειται για κύκλο και $a_i = \frac{\nu_i}{\nu} \cdot 180^0 = f_i \cdot 180^0$ αν πρόκειται για ημικύκλιο.

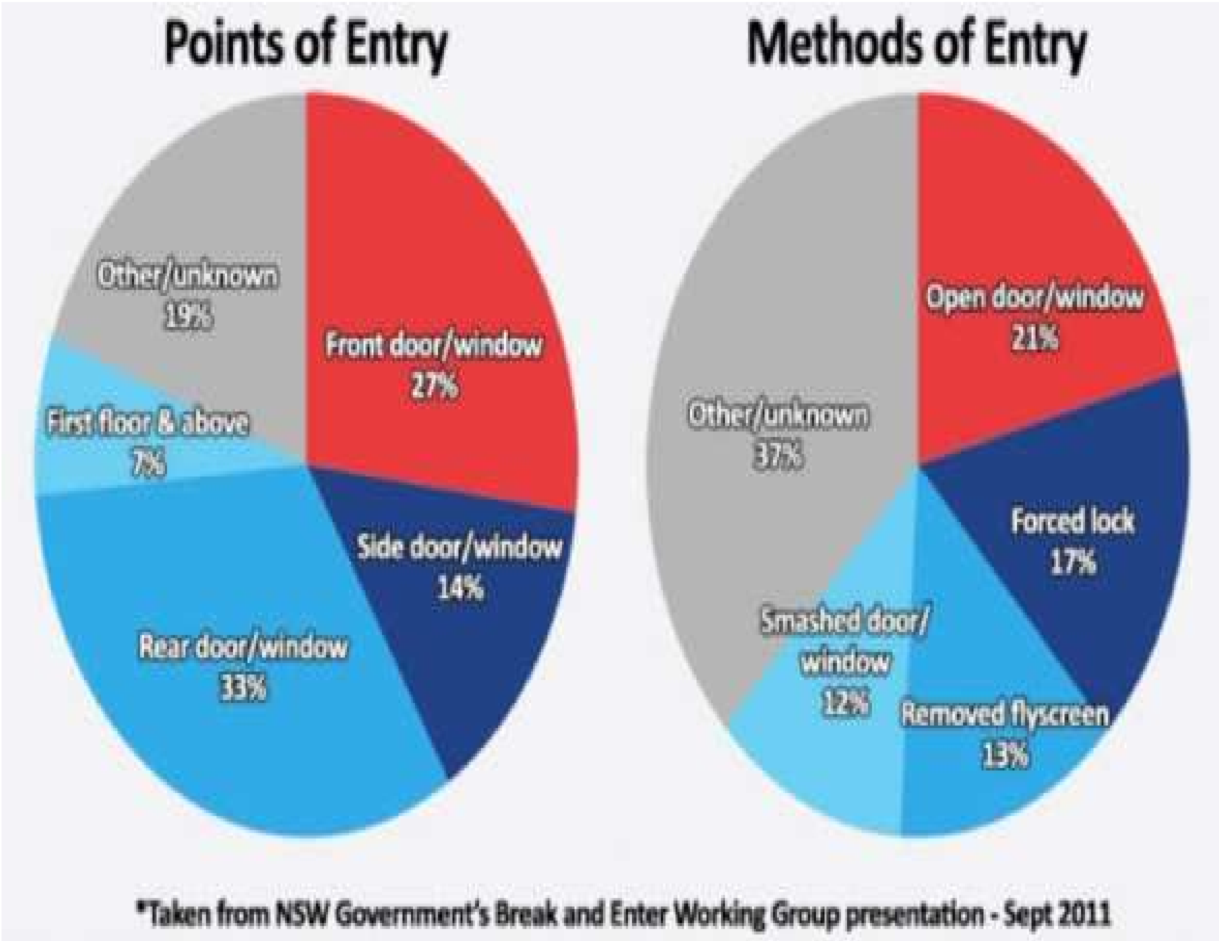


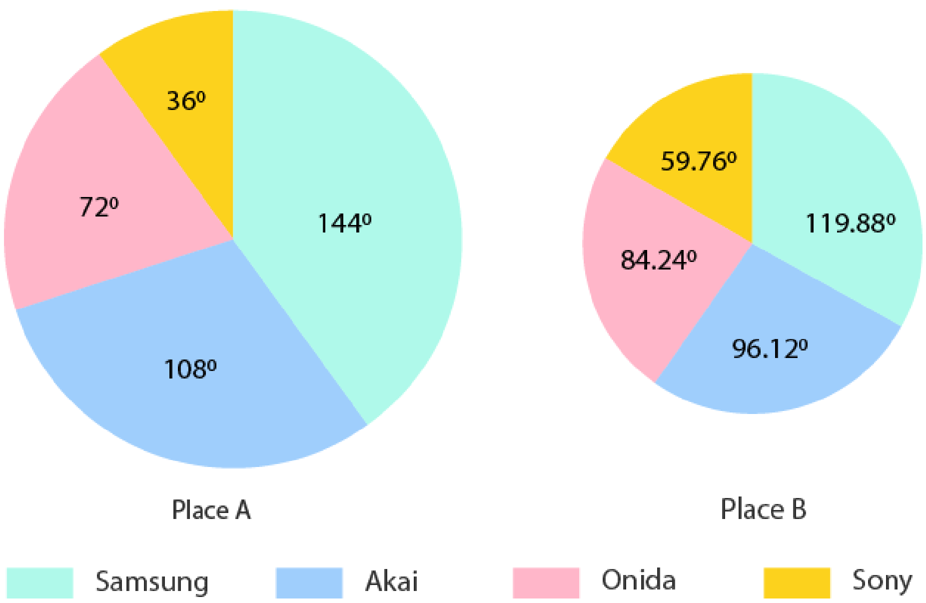









Points of Entry
Other/unknown 19%
Front door/window 27%
First floor & above 7%
Side door/window 14%
Rear door/window 33%
Methods of Entry
Open door/window 21%
Other/unknown 37%
Forced lock 17%
Smashed door/ window 12%
Removed flyscreen 13%
*Taken from NSW Government's Break and Enter Working Group presentation - Sept 2011


36⁰
72⁰
144⁰
108⁰
Place A
59.76⁰
119.88⁰
84.24⁰
96.12⁰
Place B
Samsung
Akai
Onida
Sony

Pie diagram

Μέτρα θέσεως & μέτρα διασποράς (ή μεταβλητότητας).

Για την περαιτέρω επεξεργασία των δεδομένων χρειάζονται ορισμένα χαρακτηριστικά μέτρα που καλούνται παράμετροι της κατανομής του πληθυσμού. Οι παράμετροι είναι αριθμητικές εκφράσεις που καθορίζουν τη θέση, τη διασπορά και τη μορφή της κατανομής.

Τα μέτρα θέσεως (**location measures**) μας πληροφορούν για την θέση γύρω από την οποία είναι συγκεντρωμένες οι περισσότερες παρατηρήσεις ενώ τα μέτρα διασποράς (**measures of variability**) ή μεταβλητότητας (**dispersion measures**) μας πληροφορούν σχετικά με το πόσο συγκεντρωμένες ή διασκορπισμένες είναι οι παρατηρήσεις γύρω από τα μέτρα θέσεως. Ανάλογα αν αναφερόμαστε σε όλο τον πληθυσμό ή σε ένα δείγμα του, ισχύουν τα παρακάτω σύμβολα:

	Πληθυσμός (N)	Δείγμα (n)
Μέση τιμή	μ	$\bar{x}$
Διασπορά	σ^2	s^2

Μέση τιμή ή μέσος όρος ή αριθμητικός μέσος (average, arithmetic mean).

Μέση τιμή $\bar{x}$ ονομάζετε το άθροισμα των παρατηρήσεων, δια του πλήθους τους.

• Όταν σε δείγμα μεγέθους ν οι παρατηρήσεις μίας μεταβλητής Χ είναι $t_1, t_2, t_3, \ldots, t_\nu$, τότε η μέση τιμή δίνεται από την σχέση: $\bar{x} = \dfrac{t_1 + t_2 + t_3 + \ldots + t_\nu}{\nu} = \dfrac{\sum_{i=1}^{\nu} t_i}{\nu}$.

Ο ανωτέρω τύπος χρησιμοποιείται όταν οι παρατηρήσεις είναι λίγες και διαφορετικές μεταξύ τους. Στο άθροισμα $\sum_{i=1}^{\nu} t_i$ το ν είναι ίσο με το μέγεθος του δείγματος.

• Όταν $x_1, x_2, x_3, \ldots, x_\kappa$ είναι οι τιμές της μεταβλητής Χ και $\nu_1, \nu_2, \nu_3, \ldots, \nu_\kappa$ είναι οι αντίστοιχες συχνότητες, τότε η μέση τιμή δίνεται από τη σχέση:

$$\bar{x} = \frac{\nu_1 \cdot x_1 + \nu_2 \cdot x_2 + \nu_3 \cdot x_3 + \ldots + \nu_\kappa \cdot x_\kappa}{\nu_1 + \nu_2 + \nu_3 + \ldots + \nu_\kappa} = \frac{\sum_{i=1}^{\kappa} \nu_i \cdot x_i}{\nu}.$$

Στο άθροισμα $\sum_{i=1}^{\kappa} \nu_i \cdot x_i$ το κ δίνει το πλήθος των δυνατών τιμών της μεταβλητής και όχι το μέγεθος του δείγματος.

• Όταν $x_1, x_2, x_3, \ldots, x_\kappa$ είναι οι τιμές της μεταβλητής Χ και $f_1, f_2, f_3, \ldots, f_\kappa$ οι αντίστοιχες σχετικές συχνότητες, τότε:

$$\bar{x} = x_1 \cdot f_1 + x_2 \cdot f_2 + x_3 \cdot f_3 + \ldots + x_\kappa \cdot f_\kappa = \sum_{i=1}^{\kappa} x_i \cdot f_i .$$

Όταν γνωρίζομε τις σχετικές συχνότητες επί τοις εκατό $(f_i\%)$, χρησιμοποιούμε τον τύπο: $\bar{x} = \dfrac{x_1 \cdot f_1\% + x_2 \cdot f_2\% + x_3 \cdot f_3\% + \ldots + x_\kappa \cdot f_\kappa\%}{100} = \dfrac{1}{100} \cdot \sum_{i=1}^{\kappa} x_i \cdot f_i\%$.

Πλεονεκτήματα μέσης τιμής	Μειονεκτήματα μέσης τιμής
Χρησιμοποιούνται όλες οι τιμές της για τον υπολογισμό της.	Επηρεάζεται πολύ από ακραίες τιμές.
Είναι μοναδική για κάθε σύνολο δεδομένων.	Μπορεί να μην αντιστοιχεί σε κάποια δυνατή τιμή της μεταβλητής.
Ο υπολογισμός της είναι σχετικά εύκολος.	Όταν η μεταβλητή Χ είναι διακριτή τότε η μέση τιμή μπορεί να μην είναι ακέραιος.
Είναι εύκολα κατανοητή.	Δεν υπολογίζεται για ποιοτικά δεδομένα.
Χρησιμοποιείται σε μεγάλο βαθμό και για περαιτέρω στατιστική ανάλυση.	Είναι δύσκολος ο υπολογισμός της σε ομαδοποιημένα δεδομένα με ανοικτές τις ακραίες τιμές.

Ιδιότητες μέσης τιμής.

1. Αν $t_{\min}$ είναι η μικρότερη και $t_{\max}$ η μεγαλύτερη παρατήρηση, τότε $t_{\min} < \overline{x} < t_{\max}$.

2. Το άθροισμα των διαφορών των παρατηρήσεων $t_1, t_2, t_3, \ldots, t_\nu$ από την μέση τιμή $\overline{x}$ είναι πάντοτε μηδέν, δηλαδή: $\sum_{i=1}^{\nu}\left(t_i - \overline{x}\right) = 0$.

3. Αν οι παρατηρήσεις $t_1, t_2, t_3, \ldots, t_\nu$ έχουν μέση τιμή $\overline{x}$ τότε οι παρατηρήσεις $t_1 + c,\ t_2 + c,\ t_3 + c,\ \ldots,\ t_\nu + c$ έχουν μέση τιμή $\overline{x} + c$. Έτσι όταν $c > 0$ η μέση τιμή αυξάνεται ενώ όταν $c < 0$ μειώνεται.

4. Αν οι παρατηρήσεις $t_1, t_2, t_3, \ldots, t_\nu$ έχουν μέση τιμή $\overline{x}$ τότε οι παρατηρήσεις $k \cdot t_1,\ k \cdot t_2,\ k \cdot t_3,\ \ldots,\ k \cdot t_\nu$ έχουν μέση τιμή $k \cdot \overline{x}$.

5. Αν οι παρατηρήσεις $t_1, t_2, t_3, \ldots, t_\nu$ έχουν μέση τιμή $\overline{x}$ τότε οι παρατηρήσεις $k \cdot t_1 + c,\ k \cdot t_2 + c,\ k \cdot t_3 + c,\ \ldots,\ k \cdot t_\nu + c$ έχουν μέση τιμή $k \cdot \overline{x} + c$.
Συνοπτικά ισχύει ότι:

Παρατηρήσεις	Μέση τιμή
$t_1, t_2, t_3, \ldots, t_\nu$	$\overline{x}$
$t_1 + c,\ t_2 + c,\ t_3 + c,\ \ldots,\ t_\nu + c$	$\overline{x} + c$
$k \cdot t_1,\ k \cdot t_2,\ k \cdot t_3,\ \ldots,\ k \cdot t_\nu$	$k \cdot \overline{x}$
$k \cdot t_1 + c,\ k \cdot t_2 + c,\ k \cdot t_3 + c,\ \ldots,\ k \cdot t_\nu + c$	$k \cdot \overline{x} + c$

6. Αν το αρχικό δείγμα έχει χωρισθεί σε επιμέρους ομάδες και έχει υπολογισθεί η μέση τιμή της κάθε ομάδας χωριστά, η μέση τιμή του αρχικού δείγματος είναι:

$\overline{x}=\frac{\nu_1\cdot\overline{x_1}+\nu_2\cdot\overline{x_2}+\ldots+\nu_\kappa\cdot\overline{x_\kappa}}{\nu_1+\nu_2+\ldots+\nu_\kappa}$, όπου $\overline{x_1},\overline{x_2},\ldots,\overline{x_\kappa}$ οι μέσες τιμές των επιμέρους ομάδων και $\nu_1,\nu_2,\ldots,\nu_\kappa$ τα μεγέθη των επιμέρους ομάδων.

Εφαρμογή 4. Μία τάξη αποτελείται από 15 αγόρια και 25 κορίτσια και η μέση βαθμολογία των αγοριών σε ένα διαγώνισμα είναι 16 ενώ των κοριτσιών είναι 14. Η μέση βαθμολογία της τάξεως είναι: $\overline{x}=\frac{15\cdot16+25\cdot14}{15+25}=\frac{240+350}{40}=14,7$.

Εφαρμογή 5. Αν $\overline{x}=5$ υπολογίστε τις τιμές των $\alpha,\beta\in\mathbb{R}$ του παρακάτω στατιστικού πίνακα.

x_i	f_i	$x_i\cdot f_i$
2	0,1	
3	α	
5	0,4	
6	β	
8	0,2	
Σύνολα	$\sum f_i=$	$\sum x_i\cdot f_i=$

Λύση.

Ισχύει ότι $\sum_{i=1}^{5} f_i=1\Leftrightarrow 0,1+a+0,4+\beta+0,2=1\Leftrightarrow \alpha+\beta+0,7=1\Leftrightarrow \alpha+\beta=0,3$

Ισχύει ότι $\overline{x}=5\Leftrightarrow\sum_{i=1}^{5}x_if_i=5\Leftrightarrow x_1f_1+x_2f_2+x_3f_3+x_4f_4+x_5f_5=5\Leftrightarrow$

$$2\frac{1}{10}+3a+5\frac{4}{10}+6\beta+8\frac{2}{10}=5\Leftrightarrow 3\alpha+6\beta+3,8=5\Leftrightarrow 3\alpha+6\beta=1,2\Leftrightarrow \alpha+2\beta=0,4$$

Συνεπώς είναι $\left\{\begin{matrix}\alpha+2\beta=0,4\\ \alpha+\beta=0,3\end{matrix}\right\}\Leftrightarrow\left\{\begin{matrix}\alpha=0,2\\ \beta=0,1\end{matrix}\right.$.

Εφαρμογή 6. Συμπληρώστε τον παρακάτω πίνακα που παρουσιάζει τη βαθμολογία εικοσιπέντε (25) σπουδαστών στο μάθημα «Στατιστική». Ποια η μέση βαθμολογία; Ποιο είναι το ποσοστό των σπουδαστών που προάγονται;

Βαθμολογία	Απόλυτη συχνότητα ν_i	Σχετική επί % συχνότητα $f_i\%$	Σχετική συχνότητα f_i	Κέντρο κλάσεως x_i	$x_i\nu_i$	Αθροιστική συχνότητα N_i	$x_i f_i$
$[0,\ 2)$	$\nu_1 = 3$	$f_1\% =$	$f_1 =$	$x_1 =$	$x_1\nu_1 =$	$N_1 =$	$x_1 f_1 =$
$[2,\ 4)$	$\nu_2 = 5$	$f_2\% =$	$f_2 =$	$x_2 = 3$	$x_2\nu_2 = 15$	$N_2 =$	$x_2 f_2 =$
$[4,\ 6)$	$\nu_3 = 7$	$f_3\% =$	$f_3 =$	$x_3 = 5$	$x_3\nu_3 = 35$	$N_3 =$	$x_3 f_3 =$
$[6,\ 8)$	$\nu_4 = 4$	$f_4\% =$	$f_4 =$	$x_4 =$	$x_4\nu_4 =$	$N_4 =$	$x_4 f_4 =$
$[8,\ 10]$	$\nu_5 = 6$	$f_5\% =$	$f_5 =$	$x_5 =$	$x_5\nu_5 =$	$N_5 =$	$x_5 f_5 =$
Σύνολο	$\nu = \sum_{i=1}^{5} \nu_i = 25$	$\sum_{i=1}^{5} f_i\% =$	$\sum_{i=1}^{5} f_i =$		$\sum_{i=1}^{5} x_i\nu_i =$		$\sum_{i=1}^{5} x_i f_i =$

Λύση.

Η μέση βαθμολογία προκύπτει από τον τύπο $\bar{x} = \dfrac{\sum_{i=1}^{5} x_i\nu_i}{\nu} = \dfrac{135}{25} = \dfrac{27}{5} = 5\dfrac{2}{5} = 5,4$.

Ισχύει ότι $\bar{x} = \dfrac{\sum_{i=1}^{5} x_i\nu_i}{\nu} = \dfrac{x_1\nu_1 + x_2\nu_2 + \ldots + x_5\nu_5}{\nu} = \dfrac{x_1\nu_1}{\nu} + \dfrac{x_2\nu_2}{\nu} + \ldots + \dfrac{x_5\nu_5}{\nu} = x_1\dfrac{\nu_1}{\nu} + x_2\dfrac{\nu_2}{\nu} + \ldots + x_5\dfrac{\nu_5}{\nu} =$

$$x_1 f_1 + x_2 f_2 + \ldots + x_5 f_5 = \sum_{i=1}^{5} x_i f_i = 5,4$$

Βαθμολογία	Απόλυτη συχνότητα ν_i	Σχετική επί % συχνότητα $f_i\%$	Σχετική συχνότητα f_i	Κέντρο κλάσεως x_i	$x_i\nu_i$	Αθροιστική συχνότητα N_i	$x_i f_i$
$[0,\ 2)$	$\nu_1 = 3$	$f_1\% = 12\%$	$f_1 = 0{,}12$	$x_1 = 1$	$x_1\nu_1 = 3$	$N_1 = 3$	$x_1 f_1 = 0{,}12$
$[2,\ 4)$	$\nu_2 = 5$	$f_2\% = 20\%$	$f_2 = 0{,}20$	$x_2 = 3$	$x_2\nu_2 = 15$	$N_2 = 8$	$x_2 f_2 = 0{,}60$
$[4,\ 6)$	$\nu_3 = 7$	$f_3\% = 28\%$	$f_3 = 0{,}28$	$x_3 = 5$	$x_3\nu_3 = 35$	$N_3 = 15$	$x_3 f_3 = 1{,}40$
$[6,\ 8)$	$\nu_4 = 4$	$f_4\% = 16\%$	$f_4 = 0{,}16$	$x_4 = 7$	$x_4\nu_4 = 28$	$N_4 = 19$	$x_4 f_4 = 1{,}12$
$[8,\ 10)$	$\nu_5 = 6$	$f_5\% = 24\%$	$f_5 = 0{,}24$	$x_5 = 9$	$x_5\nu_5 = 54$	$N_5 = 25$	$x_5 f_5 = 2{,}16$
Σύνολο	$\nu = \sum_{i=1}^{5} \nu_i = 25$	$\sum_{i=1}^{5} f_i\% = 100\%$	$\sum_{i=1}^{5} f_i = 1$		$\sum_{i=1}^{5} x_i\nu_i = 135$		$\sum_{i=1}^{5} x_i f_i = 5{,}4$

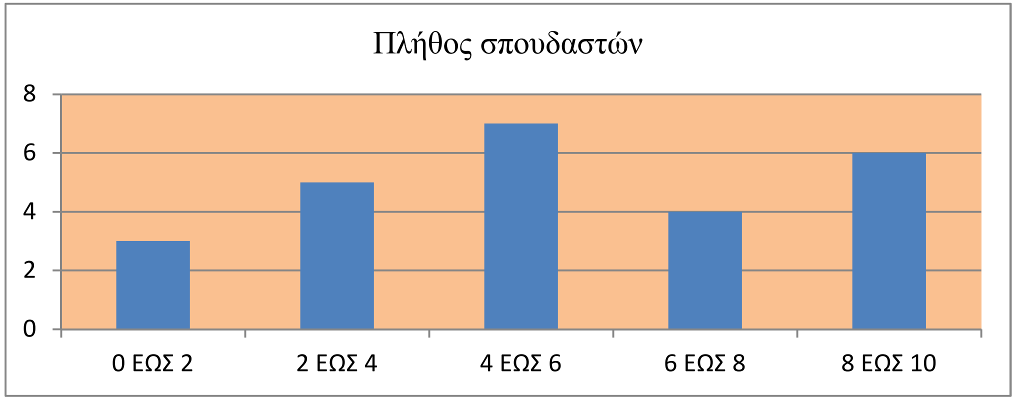
Προάγονται όσοι σπουδαστές έλαβαν βαθμό μεγαλύτερο ή ίσο του πέντε, άρα $f_3\% + f_4\% + f_5\% = 28\% + 16\% + 24\% = 68\%$.

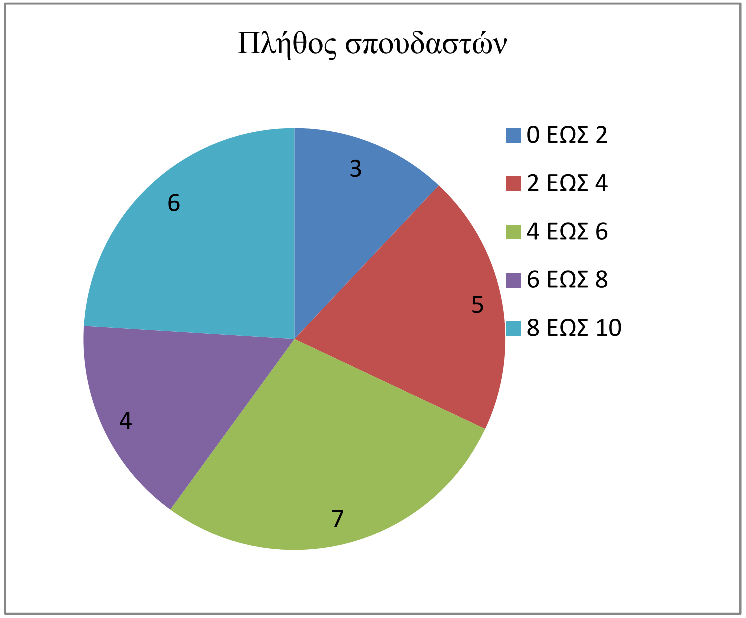
Απορρίπτονται όσοι σπουδαστές έλαβαν βαθμό μικρότερο του πέντε, άρα $f_1\% + f_2\% = 12\% + 20\% = 32\%$.

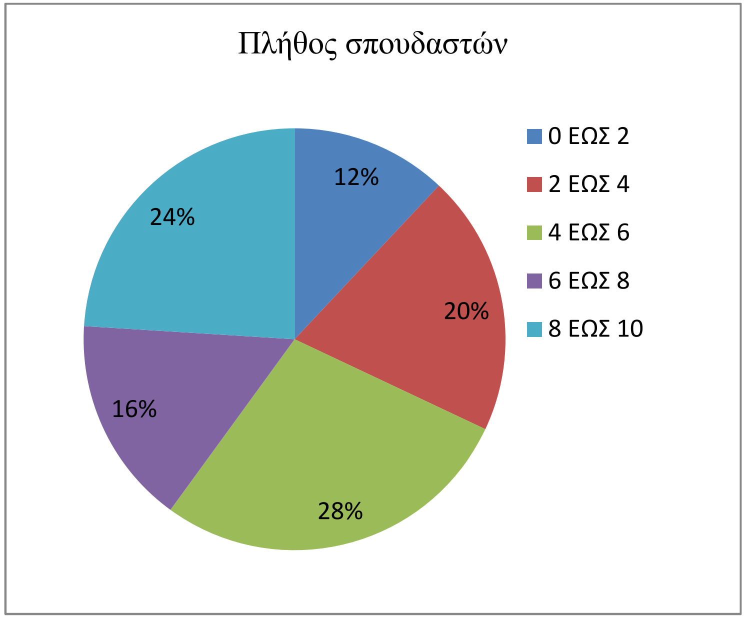
Από τους 25 σπουδαστές, οι 08 απορρίπτονται ενώ οι 17 προάγονται.

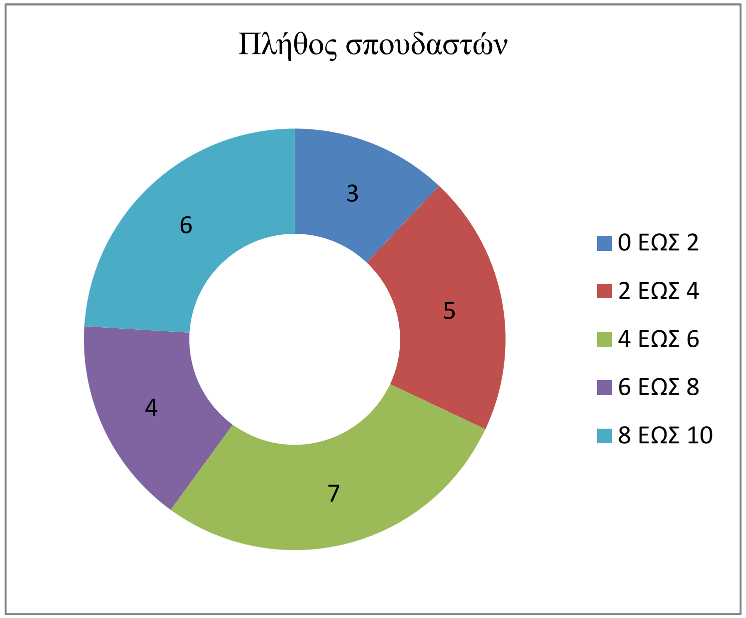
Επικρατούσα κλάση είναι αυτή με τη μεγαλύτερη συχνότητα, άρα η $[4,\ 6)$.

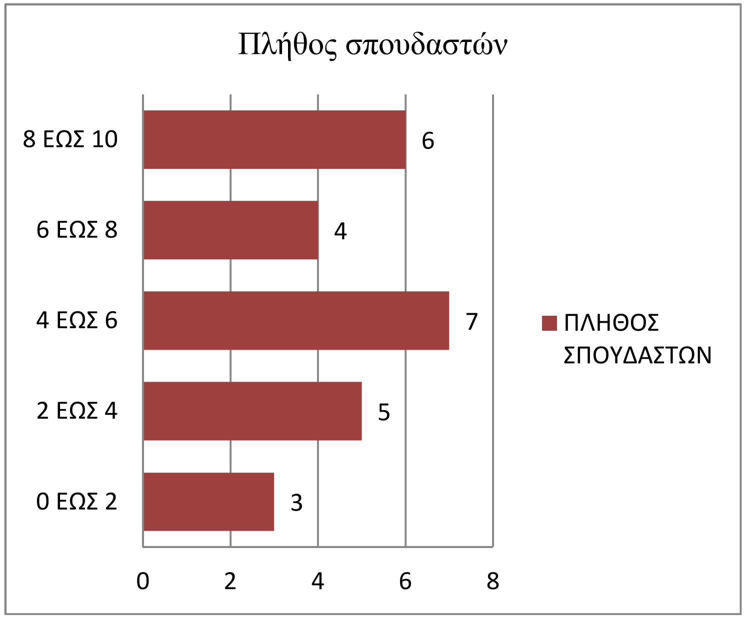
Διάμεσος είναι η $\delta = 5$ διότι είναι μεγαλύτερη από τις μισές παρατηρήσεις και μικρότερη από τις άλλες μισές.

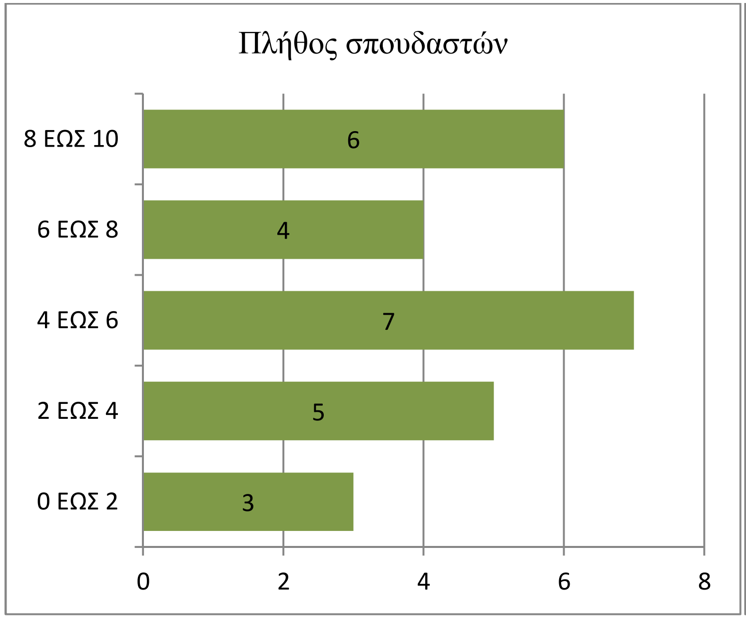


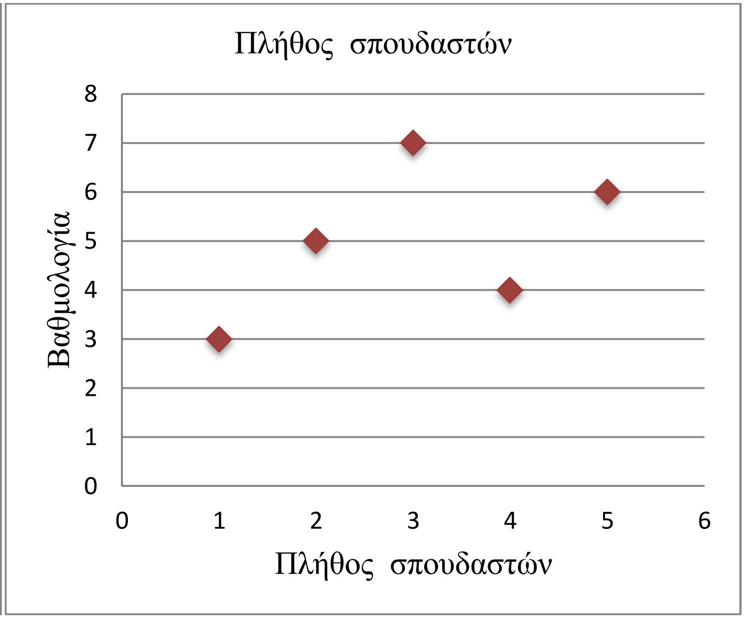

Πλήθος σπουδαστών
0 ΕΩΣ 2
2 ΕΩΣ 4
4 ΕΩΣ 6
6 ΕΩΣ 8
8 ΕΩΣ 10
3
5
7
4
6


Πλήθος σπουδαστών
0 ΕΩΣ 2
2 ΕΩΣ 4
4 ΕΩΣ 6
6 ΕΩΣ 8
8 ΕΩΣ 10
12%
20%
28%
16%
24%


Πλήθος σπουδαστών
0 ΕΩΣ 2
2 ΕΩΣ 4
4 ΕΩΣ 6
6 ΕΩΣ 8
8 ΕΩΣ 10
3
5
7
4
6


Πλήθος σπουδαστών
8 ΕΩΣ 10
6
6 ΕΩΣ 8
4
4 ΕΩΣ 6
7
2 ΕΩΣ 4
5
0 ΕΩΣ 2
3
0
2
4
6
8
ΠΛΗΘΟΣ ΣΠΟΥΔΑΣΤΩΝ


Πλήθος σπουδαστών
8 ΕΩΣ 10
6 ΕΩΣ 8
4 ΕΩΣ 6
2 ΕΩΣ 4
0 ΕΩΣ 2
6
4
7
5
3
0
2
4
6
8


Πλήθος σπουδαστών
Βαθμολογία
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
Πλήθος σπουδαστών

Εφαρμογή 7. Συμπληρώστε τον παρακάτω πίνακας που αφορά στις τιμές x_i που λαμβάνει η μεταβλητή X που περιγράφει τους βαθμούς που συγκέντρωσαν ν σκακιστικοί σύλλογοι στη διάρκεια τουρνουά. Ποια η μέση επίδοση;

x_i	ν_i	N_i	f_i	F_i	$f_i\%$	$F_i\%$	$x_i\nu_i$
$x_1=0$	$\nu_1=$	$N_1=$	$f_1=$	$F_1=$	$f_1\%=$	$F_1\%=$	$x_1\nu_1=$
$x_2=2$	$\nu_2=4$	$N_2=6$	$f_2=0,2$	$F_2=$	$f_2\%=$	$F_2\%=$	$x_2\nu_2=$
$x_3=2,5$	$\nu_3=$	$N_3=$	$f_3=$	$F_3=0,6$	$f_3\%=$	$F_3\%=$	$x_3\nu_3=$
$x_4=3$	$\nu_4=$	$N_4=$	$f_4=$	$F_4=$	$f_4\%=25$	$F_4\%=$	$x_4\nu_4=$
$x_5=6$	$\nu_5=2$	$N_5=$	$f_5=$	$F_5=$	$f_5\%=$	$F_5\%=$	$x_5\nu_5=$
$x_6=8$	$\nu_6=$	$N_6=$	$f_6=$	$F_6=$	$f_6\%=$	$F_6\%=$	$x_6\nu_6=$
Σύνολο	$\nu=\sum_{i=1}^{6}\nu_i=$		$\sum_{i=1}^{6}f_i=1$		$\sum_{i=1}^{6}f_i\%=100$		$\sum_{i=1}^{6}x_i\nu_i=$

Λύση.

x_i	ν_i	N_i	f_i	F_i	$f_i\%$	$F_i\%$	$x_i\nu_i$
$x_1=0$	$\nu_1=2$	$N_1=2$	$f_1=0,10$	$F_1=0,10$	$f_1\%=10$	$F_1\%=10$	$x_1\nu_1=0$
$x_2=2$	$\nu_2=4$	$N_2=6$	$f_2=0,20$	$F_2=0,30$	$f_2\%=20$	$F_2\%=30$	$x_2\nu_2=8$
$x_3=2,5$	$\nu_3=6$	$N_3=12$	$f_3=0,30$	$F_3=0,60$	$f_3\%=30$	$F_3\%=0,60$	$x_3\nu_3=15$
$x_4=3$	$\nu_4=5$	$N_4=17$	$f_4=0,25$	$F_4=0,85$	$f_4\%=25$	$F_4\%=0,85$	$x_4\nu_4=15$
$x_5=6$	$\nu_5=2$	$N_5=19$	$f_5=0,10$	$F_5=0,95$	$f_5\%=10$	$F_5\%=0,95$	$x_5\nu_5=12$
$x_6=8$	$\nu_6=1$	$N_6=20$	$f_6=0,05$	$F_6=1$	$f_6\%=5$	$F_6\%=100$	$x_6\nu_6=8$
Σύνολο	$\nu=\sum_{i=1}^{6}\nu_i=20$		$\sum_{i=1}^{6}f_i=1$		$\sum_{i=1}^{6}f_i\%=100$		$\sum_{i=1}^{6}x_i\nu_i=58$

Είναι $\bar{x}=\dfrac{\sum_{i=1}^{6}x_i\nu_i}{\nu}=\dfrac{58}{20}=\dfrac{29}{10}=2,9$.

Εφαρμογή 8. Υπολογίστε τα παρακάτω αθροίσματα, αν ισχύει ότι $\sum_{i=1}^{4} x_i = 100$.

- $\sum_{i=1}^{4} 3x_i = 3\sum_{i=1}^{4} x_i = 3(x_1 + x_2 + x_3 + x_4) = 3 \cdot 100 = 300$

- $\sum_{i=1}^{4} (x_i + 5) = \sum_{i=1}^{4} x_i + \sum_{i=1}^{4} 5 = (x_1 + x_2 + x_3 + x_4) + 5 \cdot 4 = 100 + 20 = 120$

- $\sum_{i=1}^{4} (3x_i + 5) = \sum_{i=1}^{4} 3x_i + \sum_{i=1}^{4} 5 = 3\sum_{i=1}^{4} x_i + 5 \cdot 4 = 300 + 20 = 320$

Ιδιότητες του συμβολισμού $\sum$.

Στη στατιστική εμφανίζονται αθροίσματα αρκετών προσθετέων. Για πρακτικούς λόγους συμβολίζονται με το σύμβολο $\sum$ όπου $\sum_{i=1}^{\nu} x_i = x_1 + x_2 + ... + x_{\nu-1} + x_\nu$.

Ισχύει ότι

- $\sum_{i=1}^{\nu} (x_i \pm y_i) = \sum_{i=1}^{\nu} x_i \pm \sum_{i=1}^{\nu} y_i$

- $\sum_{i=1}^{\nu} k = \underbrace{k + ... + k}_{\nu-\varphi o\rho\varepsilon\varsigma} = \nu k$

- $\sum_{i=1}^{\nu} kx_i = k\sum_{i=1}^{\nu} x_i$

- $\sum_{i=1}^{\nu} x_i = \sum_{i=1}^{k} x_i + \sum_{i=k+1}^{\nu} x_i$

Σταθμικός μέσος (weighted mean).

Ο σταθμικός μέσος είναι το μέτρο που χρησιμοποιούμε στις περιπτώσεις που θέλομε να δώσομε διαφορετική βαρύτητα στις τιμές ενός συνόλου δεδομένων. Αν σε κάθε τιμή $x_1, x_2, x_3, ..., x_\kappa$ δώσομε διαφορετική βαρύτητα, που εκφράζετε με τους λεγόμενους συντελεστές στάθμισης $w_1, w_2, w_3, ... w_\kappa$, τότε ο σταθμικός μέσος δίνεται από τον τύπο: $\bar{x} = \dfrac{w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + ... + w_\kappa \cdot x_\kappa}{w_1 + w_2 + w_3 + ... + w_\kappa} = \dfrac{\sum_{i=1}^{\kappa} w_\kappa \cdot x_\kappa}{\sum_{i=1}^{\kappa} w_\kappa}$.

Εφαρμογή 9. Η επίδοση σπουδαστή σε πέντε μαθήματα είναι 6, 7, 8, 9, 10.

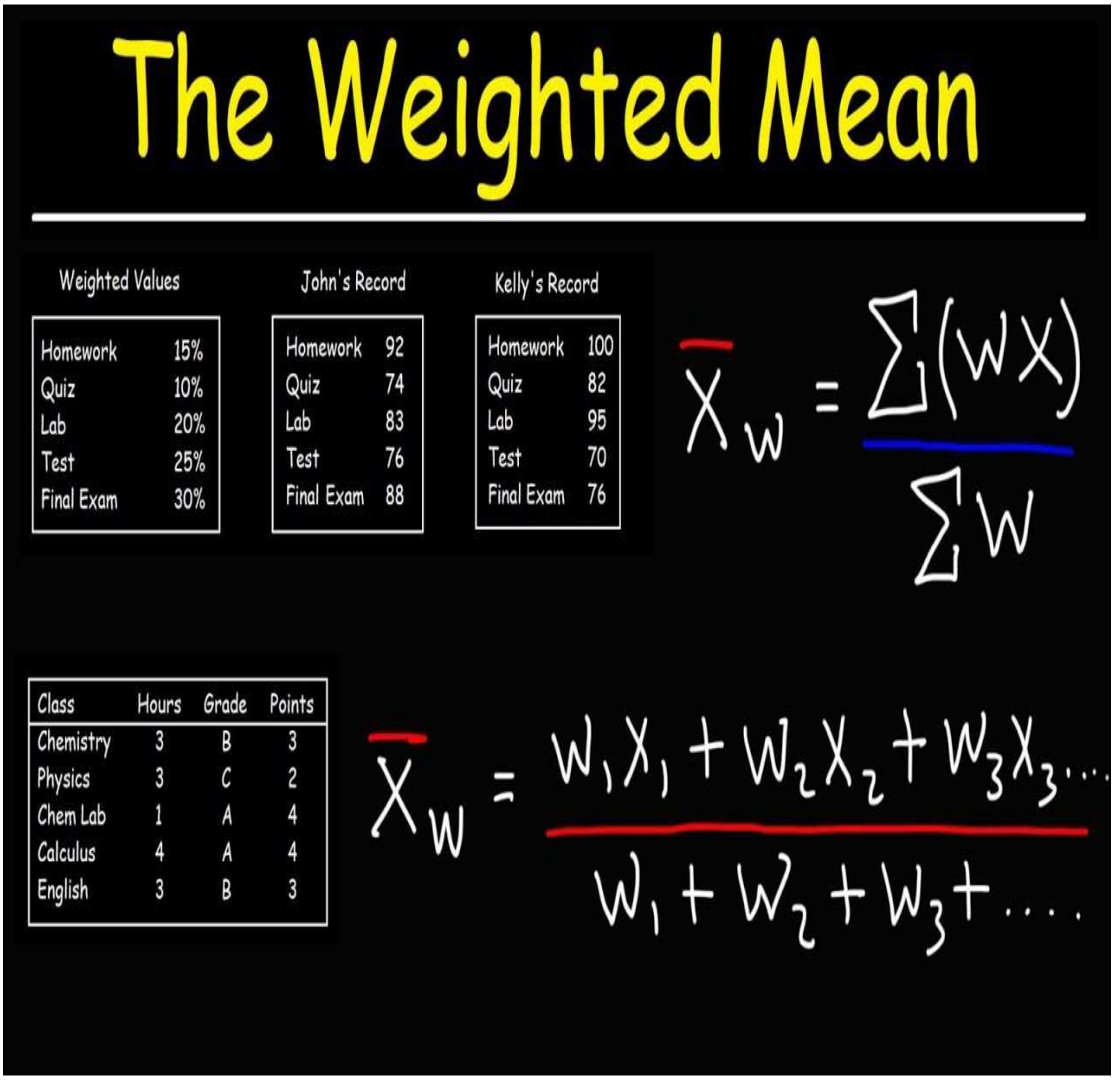
α. Ποια η μέση βαθμολογία;
β. Αν οι συντελεστές στάθμισης των μαθημάτων είναι 3, 2, 2, 2, 1 αντίστοιχα, ποια η μέση βαθμολογία;

Λύση. α. $\bar{x} = \dfrac{6+7+8+9+10}{5} = \dfrac{40}{5} = 8$

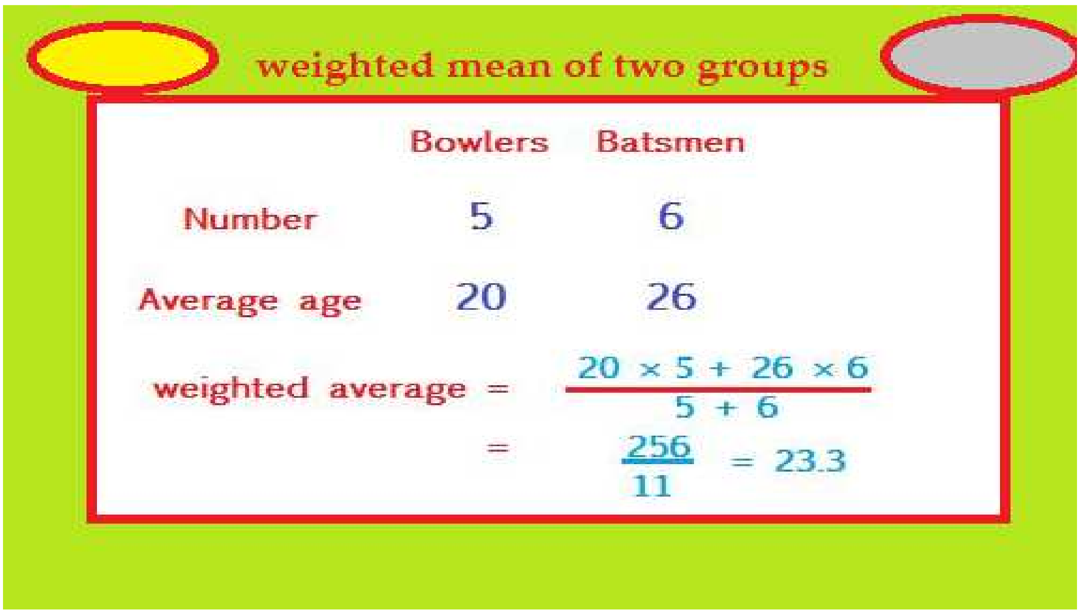
β.
$$\bar{x} = \frac{\sum_{i=1}^{5} x_i w_i}{\sum_{i=1}^{5} w_i} = \frac{x_1 w_1 + x_2 w_2 + x_3 w_3 + x_4 w_4 + x_5 w_5}{w_1 + w_2 + w_3 + w_4 + w_5} =$$
$$\frac{6 \cdot 3 + 7 \cdot 2 + 8 \cdot 2 + 9 \cdot 2 + 10 \cdot 1}{3+2+2+2+1} = \frac{76}{10} = 7{,}6$$

Εφαρμογή 10. Ο μέσος μηνιαίος μισθός των ν ναυτικών της Α εταιρείας είναι α, ενώ ο μέσος μηνιαίος μισθός των μ ναυτικών της Β εταιρείας είναι β. Ποιος θα είναι ο μέσος μηνιαίος μισθός των ναυτικών που προκύπτει από τη συγχώνευση των δυο εταιρειών; **Λύση.** $\bar{x} = \dfrac{\nu\alpha + \mu\beta}{\nu + \mu}$.

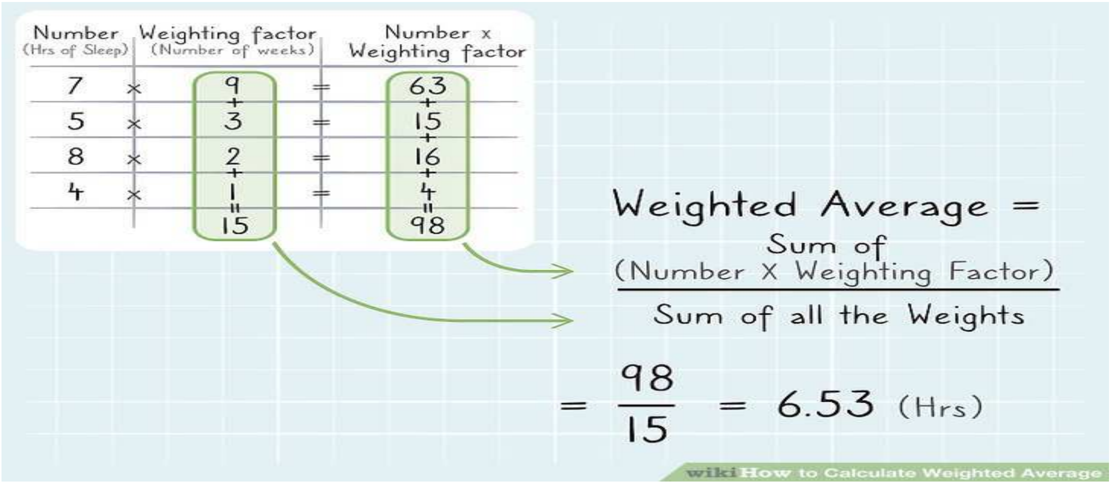
Εφαρμογή 11. Δευτεροετής σπουδαστής μελετά εβδομαδιαίως περίπου 25 ώρες. Μετά την εξεταστική περίοδο Φεβρουαρίου, αυξάνει το διάβασμα κατά 3 ώρες εβδομαδιαίως. Ποιος ο νέος αριθμητικός μέσος; **Λύση.** 4 ώρες



The Weighted Mean
Weighted Values
Homework 15%
Quiz 10%
Lab 20%
Test 25%
Final Exam 30%
John's Record
Homework 92
Quiz 74
Lab 83
Test 76
Final Exam 88
Kelly's Record
Homework 100
Quiz 82
Lab 95
Test 70
Final Exam 76
$\bar{X}_w = \frac{\sum(wx)}{\sum w}$
Class Hours Grade Points
Chemistry 3 B 3
Physics 3 C 2
Chem Lab 1 A 4
Calculus 4 A 4
English 3 B 3
$\bar{X}_w = \frac{w_1x_1 + w_2x_2 + w_3x_3 \ldots}{w_1 + w_2 + w_3 + \ldots}$

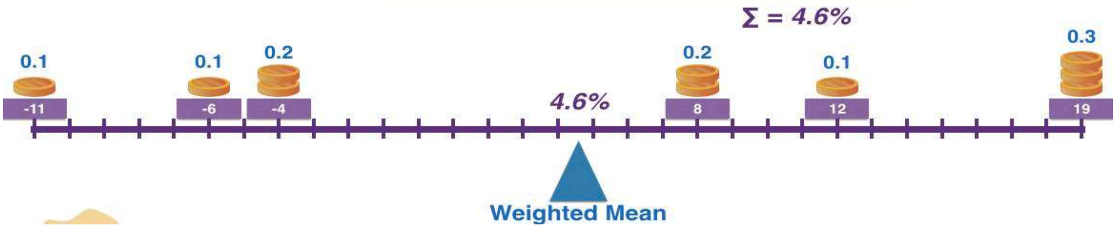


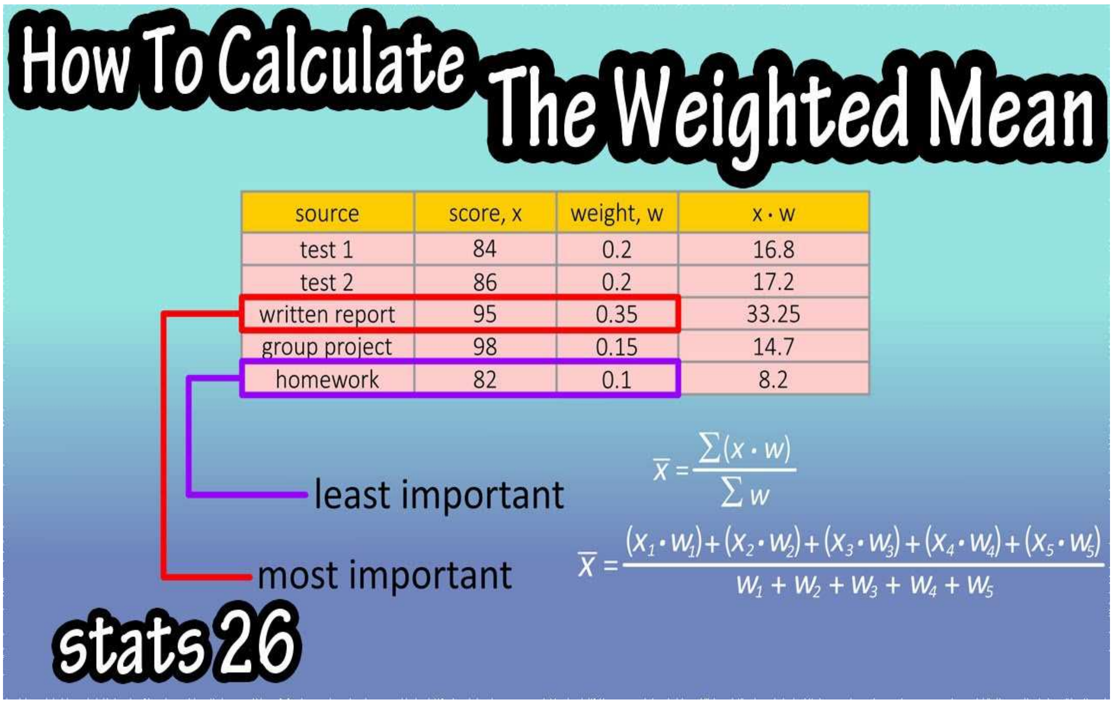
weighted mean of two groups
Bowlers Batsmen
Number 5 6
Average age 20 26
weighted average = $\frac{20 \times 5 + 26 \times 6}{5 + 6}$
$= \frac{256}{11} = 23.3$

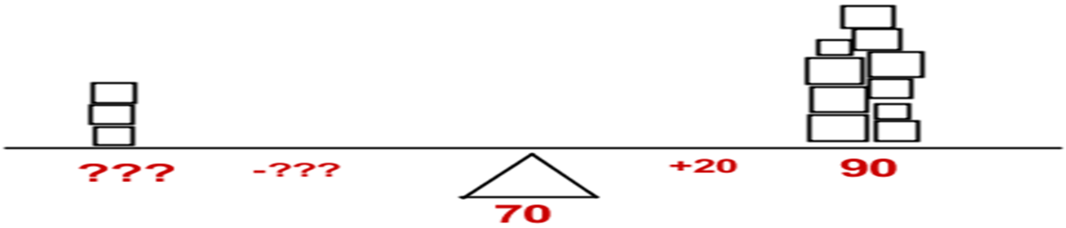


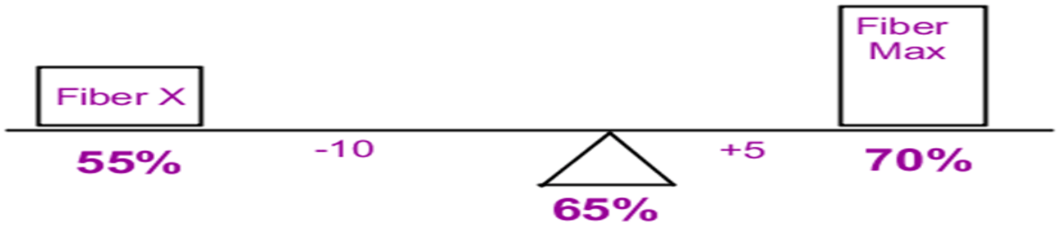
	% allocation	1yr Performance	Contribution
US Stocks	30	+12%	**+3.6%**
EU Stocks	10	-6%	**-0.6%**
Asian Stocks	10	+19%	**+1.9%**
EM Stocks	20	+8%	**+1.6%**
Treasuries	20	-4%	**-0.8%**
HY Bonds	10	-11%	**-1.1%**

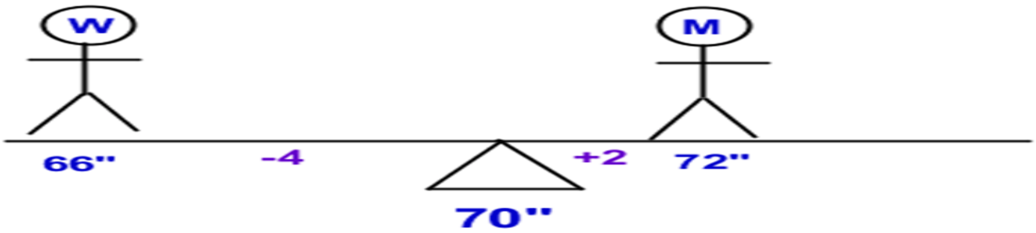
$\Sigma = 4.6\%$





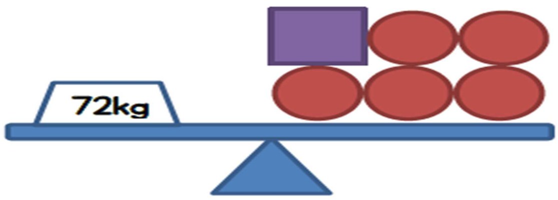






2 Here is a balance

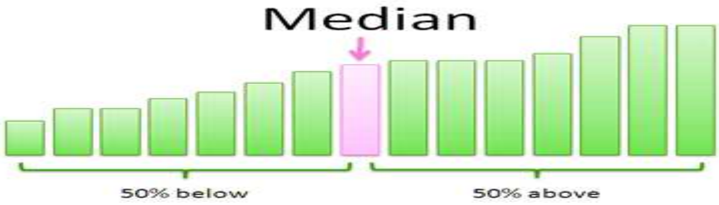
Here is another balance



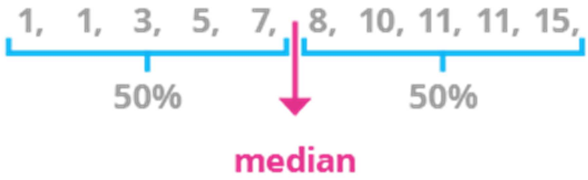
Work out the weight of **two**  s.

Διάμεσος (median).

Έστω οι ακόλουθες παρατηρήσεις: 10, 8, 7, 9, 6. Τις διατάσω κατά αύξουσα σειρά: 6, 7, 8, 9, 10. Ο μέσος είναι $\overline{x}=8$ και η διάμεσος είναι $\delta=8$.



Ορισμός διαμέσου. Η τιμή εκείνη δ ώστε το 50% των παρατηρήσεων να είναι μικρότερο ή ίσο του δ και το άλλο 50% των παρατηρήσεων μεγαλύτερο ή ίσο του δ.



Έστω οι παρατηρήσεις: 6, 7, 8, 9, 10, 11. Τότε δ είναι οποιαδήποτε τιμή μεταξύ των αριθμών 8 και 9. Δηλαδή $\delta\in(8,\ 9)$.

Αν $t_1,t_2,t_3,\ldots,t_\nu$ οι παρατηρήσεις ενός δείγματος μεγέθους ν τις οποίες έχομε διατάξει σε αύξουσα σειρά, τότε:

$$\delta=\begin{cases}\text{η μεσαια παρατηρηση, οταν ν περιττος}\\ \text{ο μεσος ορος των δυο μεσαιων παρατηρησεων, οταν ν αρτιος}\end{cases}$$

1, 3, 3, **6**, 7, 8, 9
Median = 6

1, 2, 3, **4**, **5**, 6, 8, 9
Median = (4 + 5) ÷ 2
= 4.5

• Αν το πλήθος των παρατηρήσεων είναι αρκετά μεγάλο και περιττός αριθμός, ισχύει ότι: $\delta=t_{\frac{\nu+1}{2}}$.

Εφαρμογή 12. Για τις παρατηρήσεις: $\underbrace{2,2,\ldots,2}_{39\text{ φορες}},\ \underbrace{5,5,\ldots,5}_{11\text{ φορες}},\ \underbrace{9,9,\ldots,9}_{19\text{ φορες}}$ είναι $\nu=39+11+19=69$. Άρα, $\delta=t_{\frac{\nu+1}{2}}=t_{\frac{69+1}{2}}=t_{35}=2$.

• Αν το πλήθος των παρατηρήσεων είναι αρκετά μεγάλο και άρτιος αριθμός, ισχύει ότι: $\delta=\dfrac{t_{\frac{\nu}{2}}+t_{\frac{\nu}{2}+1}}{2}$.

Εφαρμογή 13. Για τις παρατηρήσεις: $\underbrace{2,2,...,2}_{50\ \text{φορες}}$, $\underbrace{5,5,...,5}_{20\ \text{φορες}}$, $\underbrace{9,9,...,9}_{30\ \text{φορες}}$ είναι $\nu = 50+30+20=100$

Άρα, $\delta = \dfrac{t_{\frac{\nu}{2}} + t_{\frac{\nu}{2}+1}}{2} = \dfrac{t_{\frac{100}{2}} + t_{\frac{100}{2}+1}}{2} = \dfrac{t_{50}+t_{51}}{2} = \dfrac{2+5}{2} = 3,5$.

Αν έχομε ομαδοποιημένα δεδομένα, τότε επειδή η διάμεσος είναι η τιμή για την οποία το πολύ 50% των παρατηρήσεων είναι μικρότερες από αυτήν και το πολύ 50% των παρατηρήσεων είναι μεγαλύτερες από αυτήν, θα έχει αθροιστική συχνότητα (Cumulative frequency) $F_i = 50\%$.

Πλεονεκτήματα διαμέσου	**Μειονεκτήματα διαμέσου**
Είναι εύκολα κατανοητή.	Δε χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της.
Δεν επηρεάζεται από ακραίες τιμές.	Είναι δύσκολή η εφαρμογή της για περαιτέρω στατιστική ανάλυση.
Ο υπολογισμός της είναι απλός.	Δεν υπολογίζεται για ποιοτικά δεδομένα.
Είναι μοναδική σε κάθε σύνολο δεδομένων.	

Ιδιότητες διαμέσου.

1. Η διάμεσος είναι η τιμή που χωρίζει ένα σύνολο παρατηρήσεων που έχουν διαταχθεί σε αύξουσα σειρά, σε δυο ίσα μέρη. Η διάμεσος δηλαδή είναι η τιμή για την οποία το πολύ 50% των παρατηρήσεων είναι μικρότερες από αυτήν και το πολύ 50% των παρατηρήσεων είναι μεγαλύτερες από αυτήν.

2. Έστω ότι οι παρατηρήσεις $t_1, t_2, t_3, ..., t_\nu$ έχουν διάμεσο δ.

Οι παρατηρήσεις $k\cdot t_1,\ k\cdot t_2,\ k\cdot t_3,\ ...,\ k\cdot t_\nu$ έχουν διάμεσο $k\cdot\delta$.

Οι παρατηρήσεις $t_1+c,\ t_2+c,\ t_3+c,\ ...,\ t_\nu+c$ έχουν διάμεσο $\delta+c$.

Οι παρατηρήσεις $k\cdot t_1+c,\ k\cdot t_2+c,\ k\cdot t_3+c,\ ...,\ k\cdot t_\nu+c$ έχουν διάμεσο $k\cdot\delta+c$.

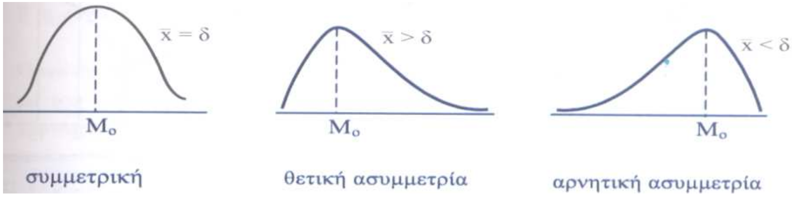
Συνοπτικά ισχύει ότι:

Παρατηρήσεις	**Διάμεσος**
$t_1, t_2, t_3, ..., t_\nu$	δ
$t_1+c,\ t_2+c,\ t_3+c,\ ...,\ t_\nu+c$	$\delta+c$
$k\cdot t_1,\ k\cdot t_2,\ k\cdot t_3,\ ...,\ k\cdot t_\nu$	$k\cdot\delta$
$k\cdot t_1+c,\ k\cdot t_2+c,\ k\cdot t_3+c,\ ...,\ k\cdot t_\nu+c$	$k\cdot\delta+c$

Αν το πλήθος των παρατηρήσεων ν είναι άρτιος αριθμός, τότε για τον υπολογισμό της διαμέσου μπορούμε αντί του ιστογράμματος (Histogram) αθροιστικών σχετικών συχνοτήτων (Cumulative relative frequencies) $F_i\%$, να έχομε ιστόγραμμα αθροιστικών συχνοτήτων N_i.

Στην περίπτωση αυτή η μόνη διαφορά είναι ότι η παράλληλη προς τον οριζόντιο άξονα δεν άγεται από το $F_i\% = 50$, αλλά από το $\frac{\nu}{2}$. Η ως άνω περιγραφόμενη διαδικασία εφαρμόζεται μόνο όταν ο ν είναι άρτιος αριθμός.

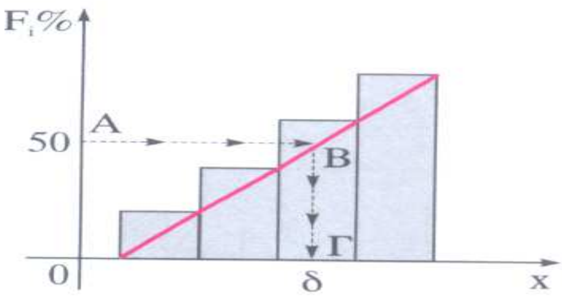
Η σχέση μέσης τιμής, διαμέσου και επικρατούσας τιμής στις διάφορες καμπύλες συχνοτήτων είναι:



Εύρεση διαμέσου δ από το πολύγωνο των αθροιστικών συχνοτήτων F_i %.

Φέρνομε από το σημείο Α (0,50) του κατακόρυφου άξονα, παράλληλη προς τον άξονα xx΄ η οποία τέμνει την πολυγωνική γραμμή στο Β.

Από το Β φέρνομε κάθετη στον άξονα Οx που τον τέμνει στο σημείο Γ. Διάμεσος δ είναι η τετμημένη του Γ.



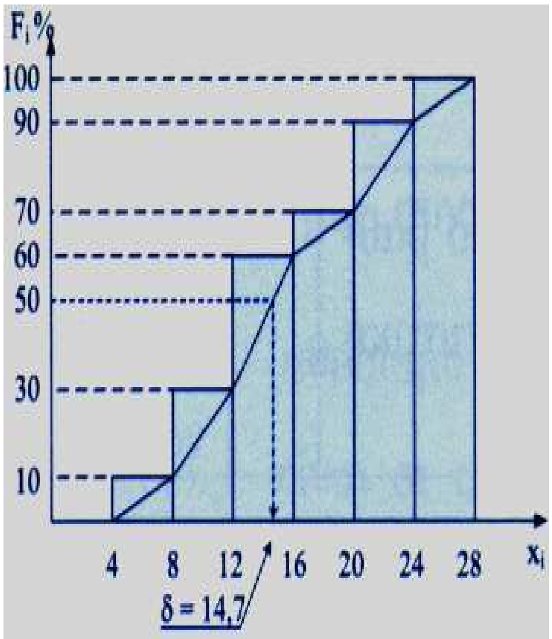
Στο διπλανό σχήμα, από το ιστόγραμμα και το πολύγωνο (**Polygon**) των σχετικών αθροιστικών % συχνοτήτων, προκύπτει η τιμή της διαμέσου δ, η οποία ως γνωστόν αντιστοιχεί στο 50% των παρατηρήσεων.

Είναι $\frac{16-\delta}{\delta-12} = \frac{60-50}{50-30} = \frac{10}{20} = \frac{1}{2}$.

Άρα,

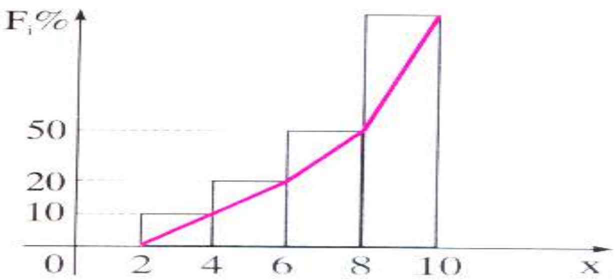
$$\frac{16-\delta}{\delta-12} = \frac{1}{2} \Leftrightarrow 2(16-\delta) = \delta - 12 \Leftrightarrow$$

$$32 - 2\delta = \delta - 12 \Leftrightarrow 44 = 3\delta \Leftrightarrow \delta = \frac{44}{3}$$

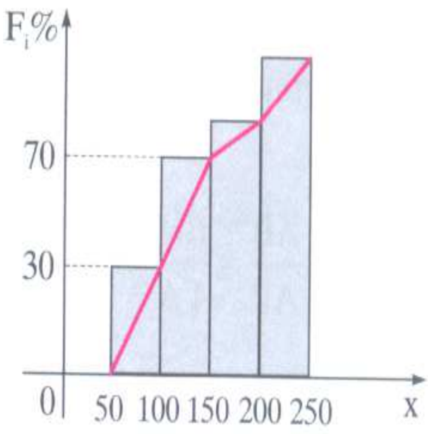


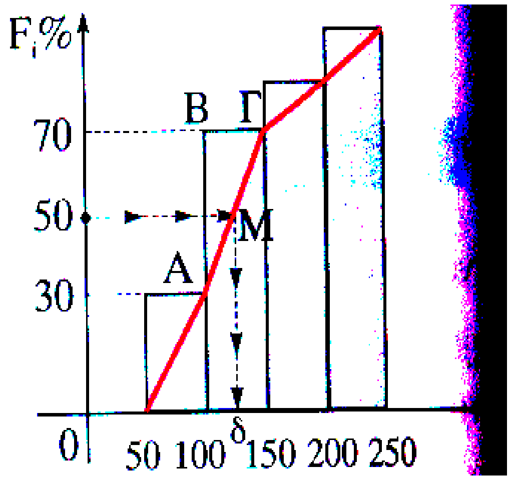
Εφαρμογή 14. Υπολογίστε τη διάμεσο δ για το παρακάτω ιστόγραμμα και πολύγωνο σχετικών αθροιστικών % συχνοτήτων.

Λύση. Το 50% των παρατηρήσεων είναι μικρότερο από το 8. Άρα δ=8.



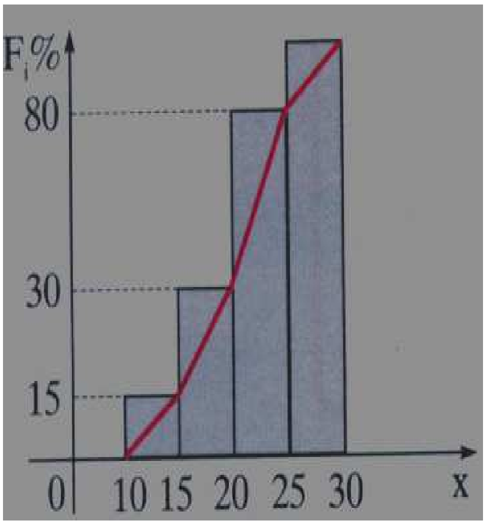
Εφαρμογή 15. Υπολογίσετε τη διάμεσο για το ανωτέρω ιστόγραμμα και πολύγωνο σχετικών αθροιστικών % συχνοτήτων.

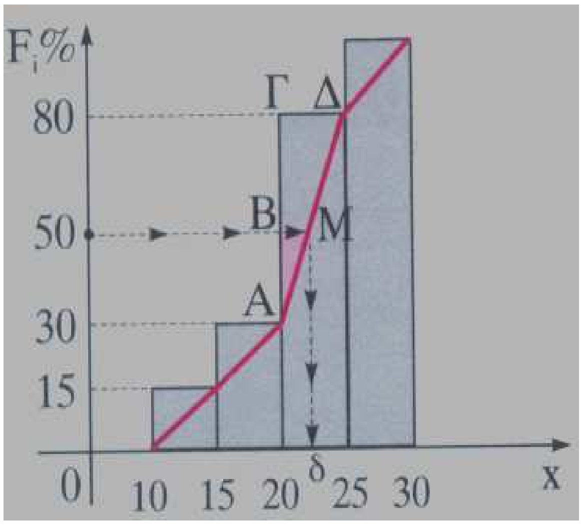




Λύση. Είναι $\frac{150-\delta}{\delta-100}=\frac{70-50}{50-30}\Leftrightarrow\frac{150-\delta}{\delta-100}=\frac{20}{20}\Leftrightarrow 150-\delta=\delta-100\Leftrightarrow\delta=125$.

Εφαρμογή 16. Υπολογίστε τη διάμεσο για το παρακάτω ιστόγραμμα και πολύγωνο σχετικών αθροιστικών % συχνοτήτων.





Λύση. Είναι $\frac{25-\delta}{\delta-20}=\frac{80-50}{50-30}=\frac{30}{20}=\frac{3}{2}$.

Άρα, $50-2\delta=3\delta-60\Leftrightarrow 110=5\delta\Leftrightarrow\delta=22$.

Εφαρμογή 17. Η μέση τιμή και η διάμεσος πέντε αριθμών είναι 4. Οι τρεις από αυτούς είναι οι 1, 2, 6. Βρείτε τους άλλους δυο αριθμούς.
Λύση. Αφού οι παρατηρήσεις είναι 5 το πλήθος, δηλαδή περιττός αριθμός, η μεσαία παρατήρηση που είναι η τρίτη, θα ισούται με την τιμή της διαμέσου, δηλαδή 4.

Διατάσσομε τους αριθμούς με αύξουσα σειρά όπως ακολούθως:

$$1 \quad 2 \quad 4 \quad 6 \quad y$$
$$\uparrow$$
$$\delta$$

Ισχύει ότι $\bar{x} = 4 \Leftrightarrow \dfrac{\sum_{i=1}^{5} x_i}{\nu} = 4 \Leftrightarrow \dfrac{1+2+4+6+y}{5} = 4 \Leftrightarrow$

$$\frac{13+y}{5} = 4 \Leftrightarrow 13 + y = 20 \Leftrightarrow y = 7.$$

Εφαρμογή 18. Στον παρακάτω στατιστικό δίνεται ο συνολικός αριθμός των εργαζομένων σε ναυτιλιακές εταιρείες του Πειραιά. Υπολογίστε τον $\alpha \in \mathbb{R}$ αν $\delta = 39$.

x_i	ν_i
$x_1 = 10$	$\nu_1 = 4$
$x_2 = 20$	$\nu_2 = 4$
$x_3 = 30$	$\nu_3 = \alpha$
$x_4 = 40$	$\nu_4 = 10$
$x_5 = 50$	$\nu_5 = 3$
Σύνολο	$\nu = \sum_{i=1}^{5} \nu_i =$

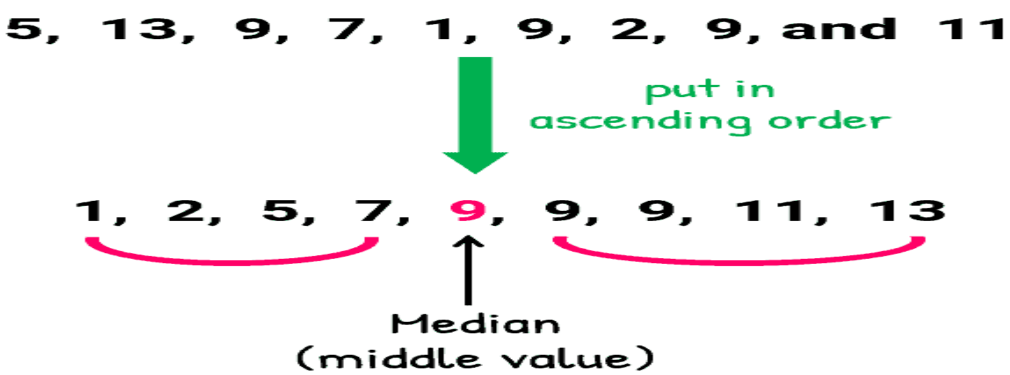
Λύση. Αφού η διάμεσος είναι 39, οι μισές παρατηρήσεις (50% των συχνοτήτων) θα είναι μικρότερες της και οι άλλες μισές μεγαλύτερες.
Ισχύει ότι 4 + 4 + α = 10 + 3. Συνεπώς α = 5.

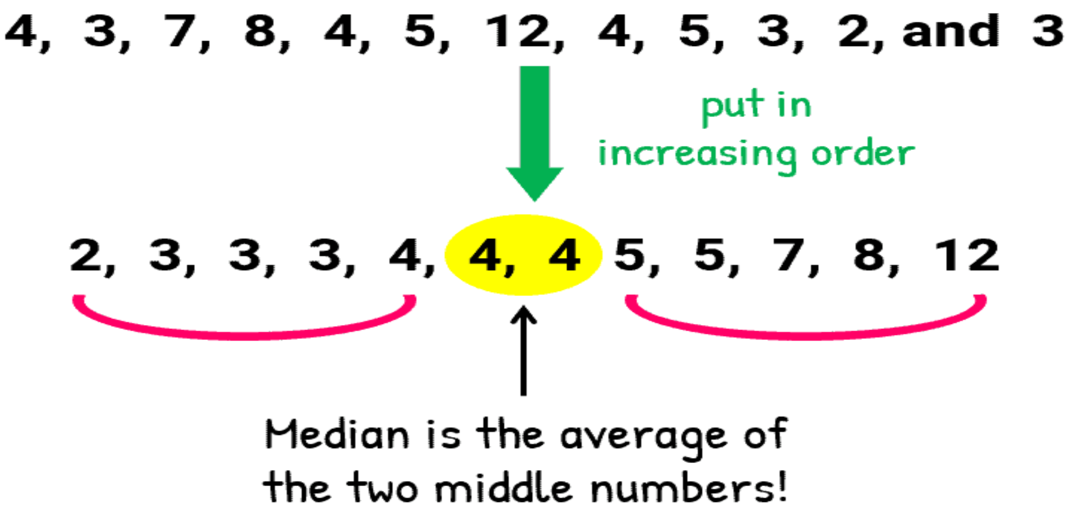
Εφαρμογή 19. Οι αριθμοί 5, 6, 10, 11, 13, x, y έχουν μέση τιμή 8 και διάμεσο 7. Υπολογίστε τους $x,\ y \in \mathbb{R}$.
Λύση. Αφού οι παρατηρήσεις είναι 7 το πλήθος, δηλαδή περιττός αριθμός, η μεσαία παρατήρηση που είναι η τέταρτη, θα ισούται με την τιμή της διαμέσου, δηλαδή 7.
Διατάσσομε τους αριθμούς με αύξουσα σειρά όπως ακολούθως:

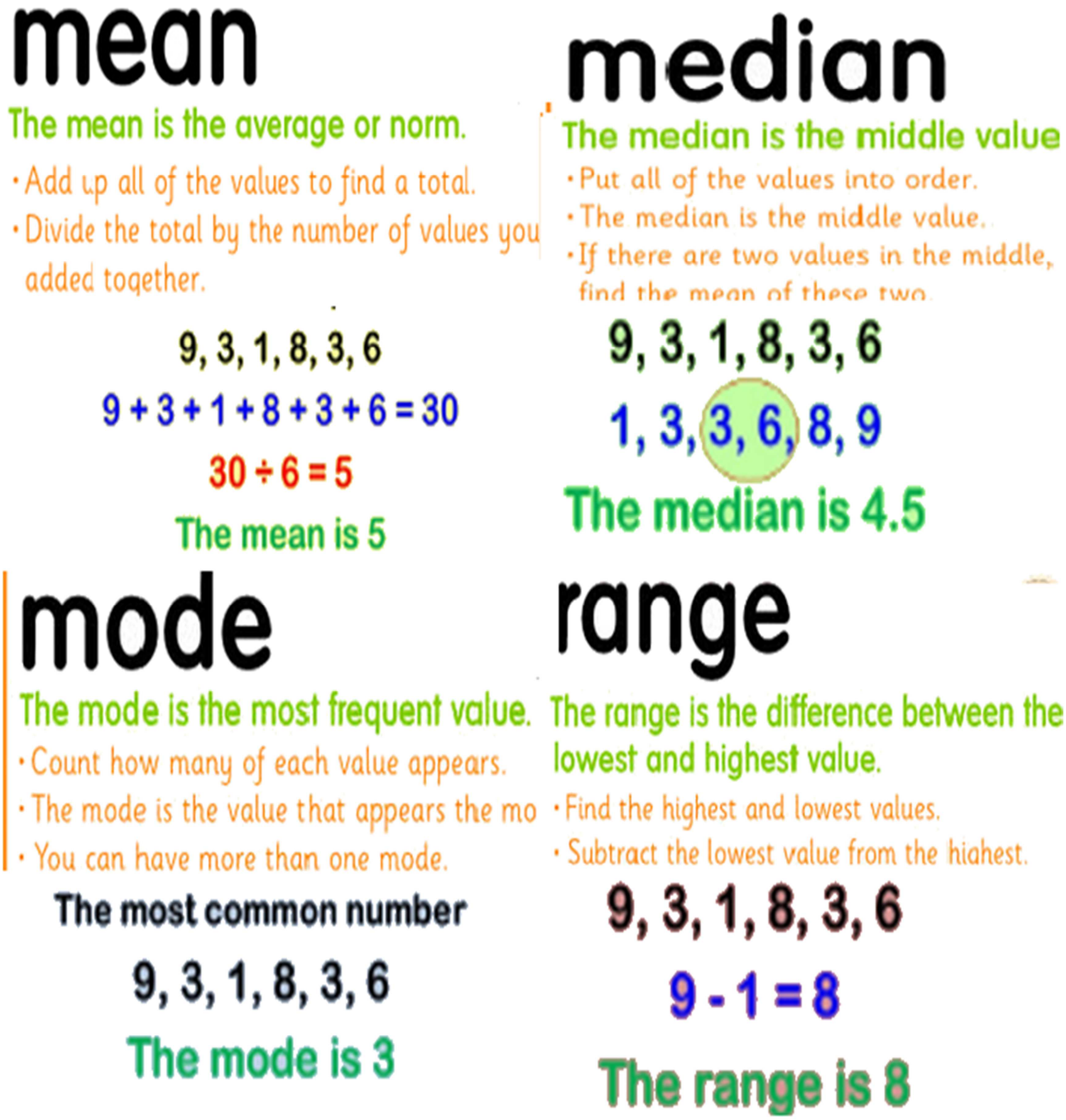
$$y \quad 5 \quad 6 \quad 7 \quad 10 \quad 11 \quad 13$$
$$\uparrow$$
$$\delta$$

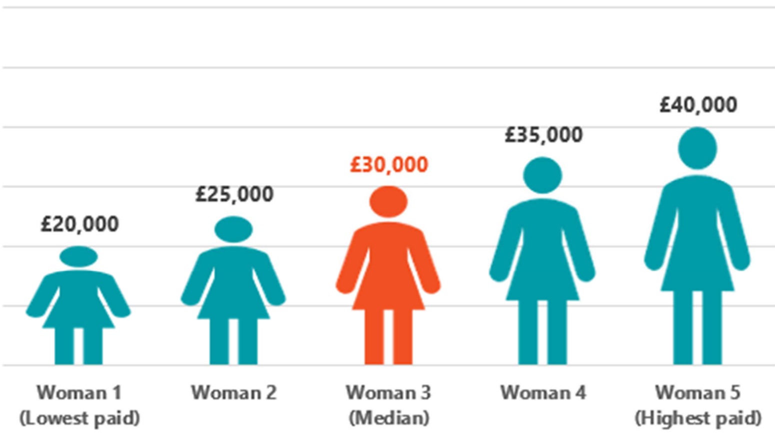
Ισχύει ότι $\bar{x} = 8 \Leftrightarrow \dfrac{\sum_{i=1}^{7} x_i}{\nu} = 4 \Leftrightarrow \dfrac{y+5+6+7+10+11+13}{7} = 8 \Leftrightarrow$

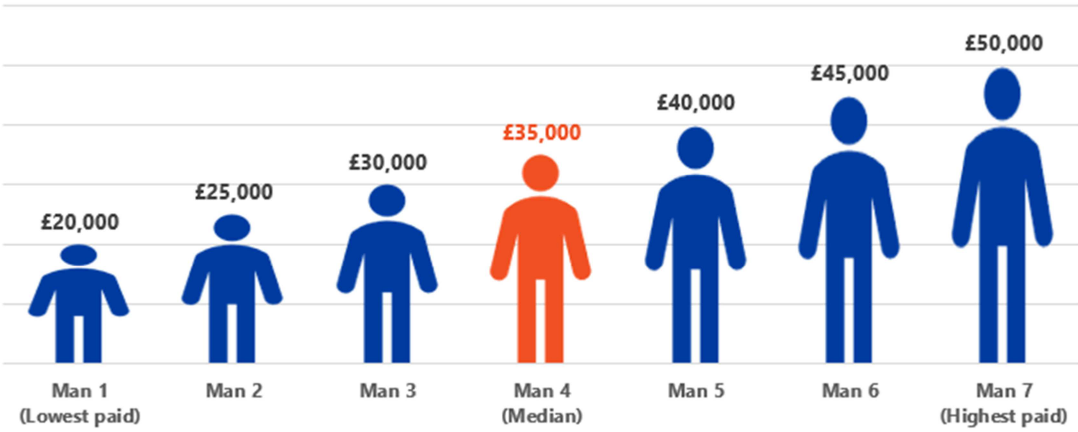
$$\frac{52+y}{7} = 8 \Leftrightarrow 52 + y = 56 \Leftrightarrow y = 4.$$

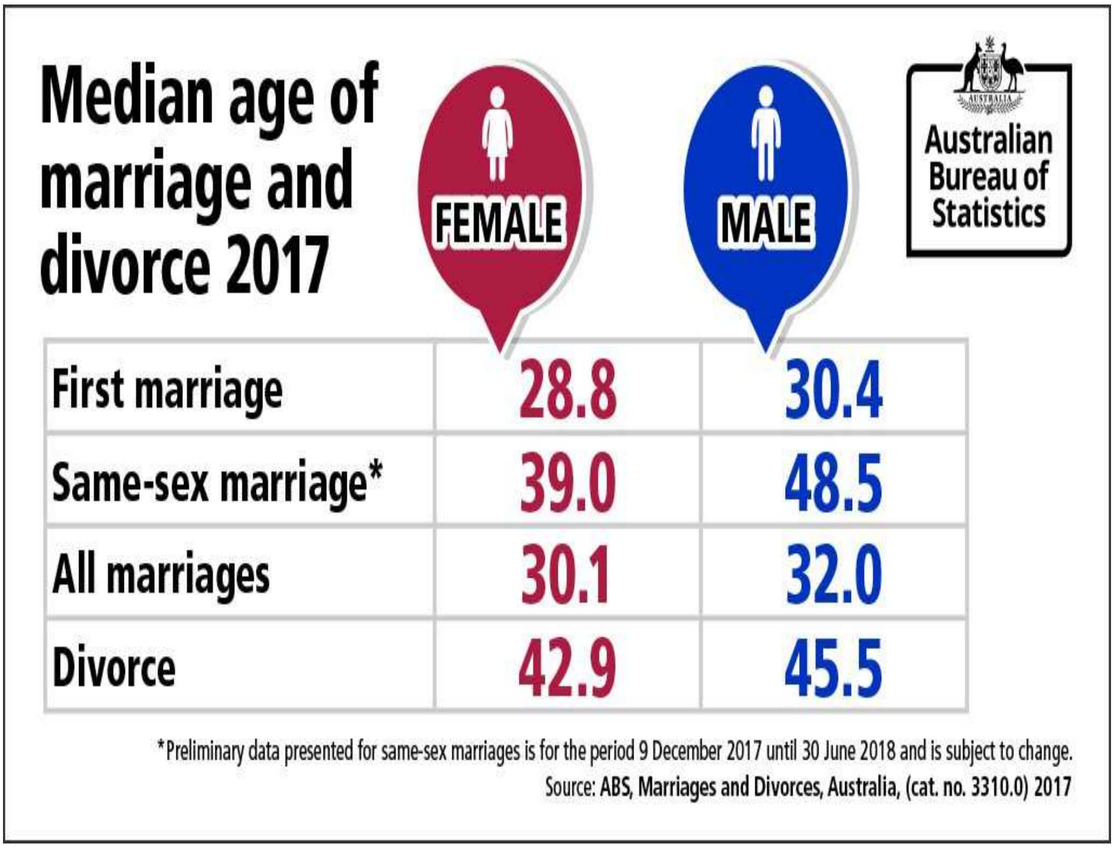



4, 3, 7, 8, 4, 5, 12, 4, 5, 3, 2, and 3
put in
increasing order
2, 3, 3, 3, 4, 4, 4 5, 5, 7, 8, 12
Median is the average of
the two middle numbers!


mean
The mean is the average or norm.
•Add up all of the values to find a total.
•Divide the total by the number of values you
added together.
9, 3, 1, 8, 3, 6
9 + 3 + 1 + 8 + 3 + 6 = 30
30 ÷ 6 = 5
The mean is 5
median
The median is the middle value
•Put all of the values into order.
•The median is the middle value.
•If there are two values in the middle,
9, 3, 1, 8, 3, 6
1, 3, 3, 6, 8, 9
The median is 4.5
mode
The mode is the most frequent value.
•Count how many of each value appears.
•You can have more than one mode.
The most common number
9, 3, 1, 8, 3, 6
The mode is 3
range
The range is the difference between the
lowest and highest value.
•Find the highest and lowest values.
•Subtract the lowest value from the highest.
9, 3, 1, 8, 3, 6
9 - 1 = 8
The range is 8







Median age of marriage and divorce 2017

	FEMALE	MALE
First marriage	28.8	30.4
Same-sex marriage*	39.0	48.5
All marriages	30.1	32.0
Divorce	42.9	45.5

*Preliminary data presented for same-sex marriages is for the period 9 December 2017 until 30 June 2018 and is subject to change.

Source: **ABS, Marriages and Divorces, Australia, (cat. no. 3310.0) 2017**

Ομαδοποίηση παρατηρήσεων.

Η ομαδοποίηση των παρατηρήσεων εφαρμόζεται όταν είναι μεγάλο το μέγεθος του δείγματος της ποσοτικής μεταβλητής X. Προς τούτο μοιράζομε τα δεδομένα σε ομάδες οι οποίες ονομάζονται κλάσεις (Class intervals), τοποθετώντας σε κάθε μία κλάση το πλήθος των παρατηρήσεων που της αντιστοιχούν.

Συνήθως, οι κλάσεις είναι διαστήματα της μορφής $[\alpha,\ \beta)$ με το α να ονομάζεται κατώτερο και το β ανώτερο όριο της κλάσεως. Τα α, β ονομάζονται όρια της κλάσεως (Class boundaries). Η κάθε κλάση χαρακτηρίζεται και διακρίνεται από τα όρια της.

- Πλάτος c της κλάσεως ονομάζεται η διαφορά του ανωτέρου από το κατώτερο όριο της. Συνεπώς ισχύει ότι $c = \beta - \alpha$.

- Κέντρο της κλάσεως ονομάζεται ο αριθμός $x = \dfrac{a+\beta}{2}$ ο οποίος αντιπροσωπεύει την κλάση, λειτουργώντας ως μία διακριτή μεταβλητή.

- Οι κεντρικοί όροι x_i των κλάσεων αποτελούν διαδοχικούς όρους αριθμητικής προόδου με πρώτο όρο x_i και διαφορά $\omega = c$. Ισχύει ότι $x_\nu = x_1 + (\nu - 1)c$.

- Ο καθορισμός του πλήθους των κλάσεων των ομαδοποιημένων παρατηρήσεων προκύπτει από τον **κανόνα του Sturges**. Αναλόγως με το μέγεθος του δείγματος και εφόσον δεν αναφέρεται διαφορετικά, στην πράξη ισχύουν τα παρακάτω:

Μέγεθος δείγματος	<20	$20-50$	$50-100$	$100-200$
Πλήθος κλάσεων k	5	6	7	8

- Για τη δημιουργία κλάσεων ίσου πλάτους, μετά την εύρεση του εύρους (Range), υπολογίζεται το πλάτος c των κλάσεων, από τον τύπο $c = \dfrac{R}{k}$. Αν το πλάτος δεν είναι ακέραιος αριθμός, γίνεται στρογγυλοποίηση του προς τα πάνω.

Στην τελευταία κλάση μόνο, είναι δυνατό το διάστημα να θεωρηθεί κλειστό και στα δυο του άκρα.

Εφαρμογή 20. Αν ο νεότερος υπάλληλος εταιρείας είναι 20 ετών και ο γηραιότερος 70 ετών, είναι $R = 70 - 20 = 50$. Αν η εταιρεία απασχολεί 19 εργαζομένους είναι $k = 5$, άρα $c = \dfrac{R}{k} = \dfrac{50}{5} = 10$. Άρα, για τη στατιστική μελέτη της ηλικιακής κατανομής τους, θα χρησιμοποιηθεί ο παρακάτω στατιστικός πίνακας.

Αριθμός κλάσεων	Ηλικία	Συχνότητα
1	$[20,\ 30)$	$\nu_1 =$
2	$[30,\ 40)$	$\nu_2 =$
3	$[40,\ 50)$	$\nu_3 =$
4	$[50,\ 60)$	$\nu_4 =$

5	$[60,\ 70)$	$\nu_5 =$
	Σύνολο	$\nu =$

Για 199 εργαζομένους είναι $k = 8$, άρα $c = \frac{R}{k} = \frac{50}{8} = \frac{25}{4} = \frac{12,50}{2} = 6,25$. Με προς τα πάνω στρογγυλοποίηση είναι $c = 6,5$. Άρα, για τη στατιστική μελέτη της ηλικιακής κατανομής τους, θα χρησιμοποιηθεί ο παρακάτω στατιστικός πίνακας.

Αριθμός κλάσεων	Ηλικία	Συχνότητα
1	$[20-26,5)$	$\nu_1 =$
2	$[26,5-33)$	$\nu_2 =$
3	$[33-39,5)$	$\nu_3 =$
4	$[39,5-46)$	$\nu_4 =$
5	$[46-52,5)$	$\nu_5 =$
6	$[52,5-59)$	$\nu_6 =$
7	$[59-65,5)$	$\nu_7 =$
8	$[65,5-72)$	$\nu_8 =$
	Σύνολο	$\nu =$

Αν το $c = 6,25$ είχε στρογγυλοποιηθεί προς τα κάτω σε $c = 6$ θα προέκυπτε ο παρακάτω στατιστικός πίνακας που είναι ατελής διότι δεν περιλαμβάνει τους άνω των 68 ετών υπαλλήλους.

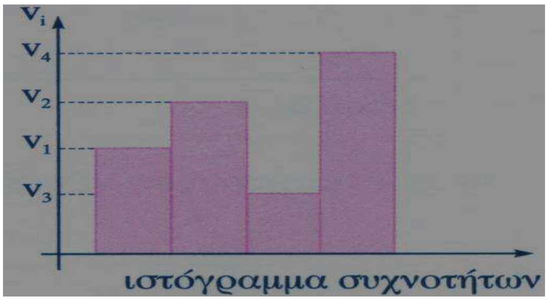
Αριθμός κλάσεων	Ηλικία	Συχνότητα
1	$[20-26)$	$\nu_1 =$
2	$[26-32)$	$\nu_2 =$
3	$[32-38)$	$\nu_3 =$
4	$[38-44)$	$\nu_4 =$
5	$[44-50)$	$\nu_5 =$
6	$[50-56)$	$\nu_6 =$
7	$[56-62)$	$\nu_7 =$
8	$[62-68)$	$\nu_8 =$
	Σύνολο	$\nu =$

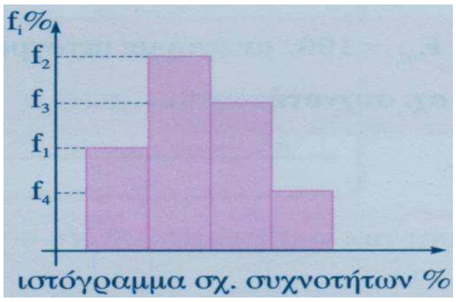
Γραφική παράσταση ομαδοποιημένης κατανομής. Ιστόγραμμα συχνοτήτων.

Το ιστόγραμμα συχνοτήτων αποτελείται από μία σειρά ορθογωνίων παραλληλογράμμων, εφαπτόμενων μεταξύ τους, που έχουν ίσες βάσεις πλάτους c και ύψη ίσα με τη συχνότητα που αντιπροσωπεύουν.

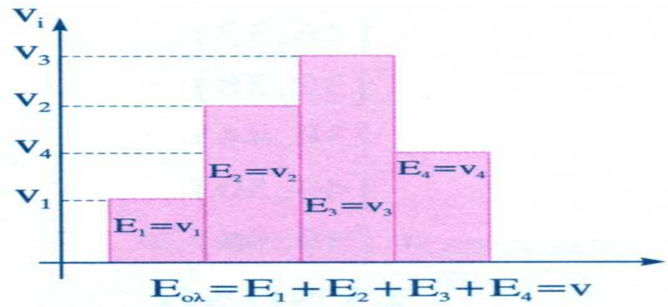
Αν το πλάτος c της κάθε κλάσεως θεωρηθεί ως μονάδα μετρήσεως του οριζοντίου άξονα, **το εμβαδό κάθε ορθογωνίου παραλληλογράμμου του**

ιστογράμματος ισούται με την αντίστοιχη συχνότητα. Για το άθροισμα των εμβαδών όλων των ορθογωνίων παραλληλογράμμου ισχύουν:

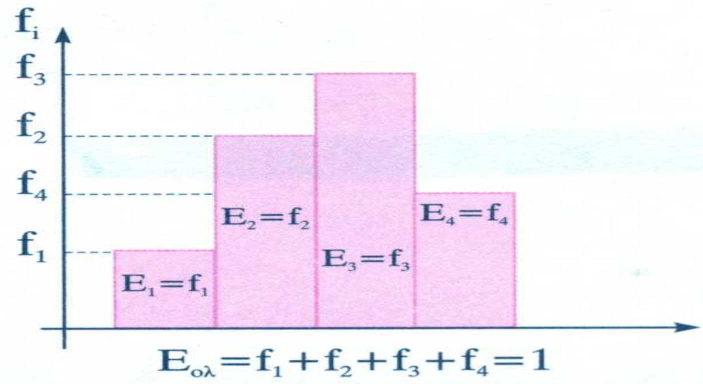




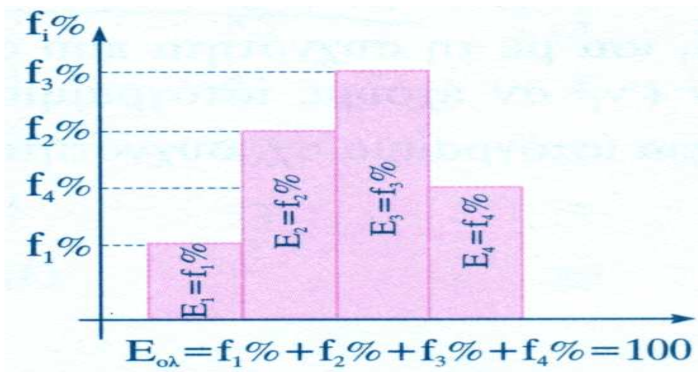
Για το ιστόγραμμα συχνοτήτων ισχύει ότι $E_{ΟΛΙΚΟ} = \nu$.



Για το ιστόγραμμα (Histogramm) των σχετικών συχνοτήτων (Relative frequencies) ισχύει ότι $E_{ΟΛΙΚΟ} = 1$.

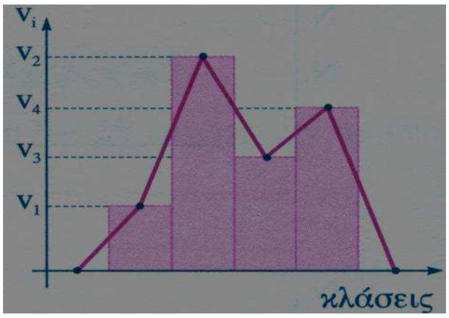


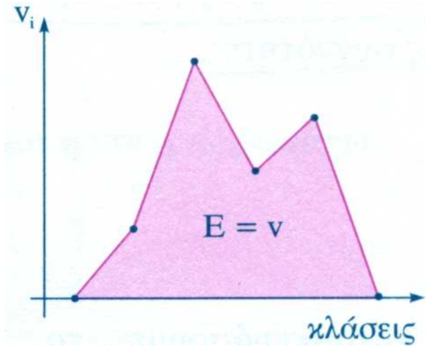
Για το ιστόγραμμα των σχετικών % συχνοτήτων ισχύει ότι $E_{ΟΛΙΚΟ} = 100$.



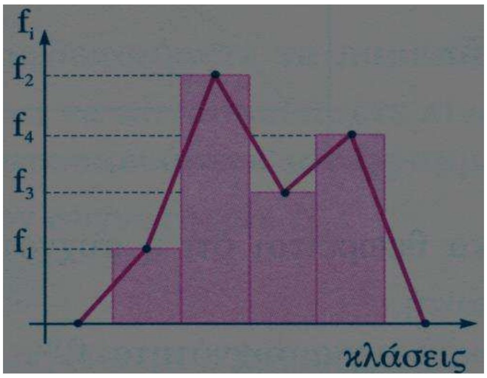
Πολύγωνο συχνοτήτων (Frequency polygon).

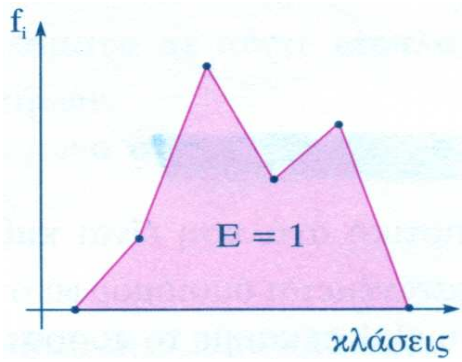
Στην αρχή και στο τέλος του ιστογράμματος θεωρούμε ότι υπάρχουν δυο υποθετικές κλάσεις, μηδενικής συχνότητας εκάστη. Το πολύγωνο συχνοτήτων είναι εκείνη η πολυγωνική γραμμή που ενώνει **τα μέσα των άνω βάσεων** των ορθογωνίων παραλληλογράμμων, περιλαμβανομένων και των υποθετικών κλάσεων.



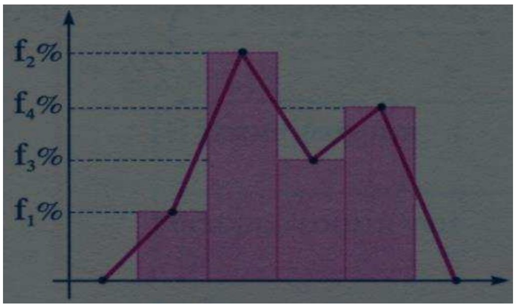


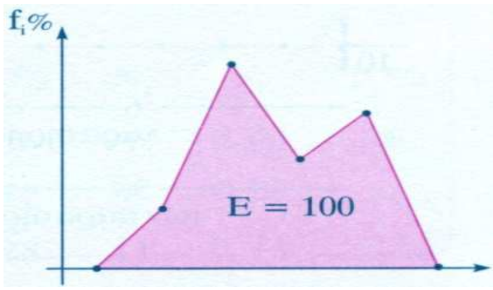
Όπως φαίνεται στο παραπάνω σχήμα, ισχύει ότι το εμβαδόν που περικλείεται από το πολύγωνο των συχνοτήτων και τον οριζόντιο άξονα ισούται με το μέγεθος του δείγματος, δηλαδή ν.





Όπως φαίνεται στο παραπάνω σχήμα, ισχύει ότι το εμβαδόν που περικλείεται από το πολύγωνο των σχετικών συχνοτήτων και τον οριζόντιο άξονα ισούται με 1.

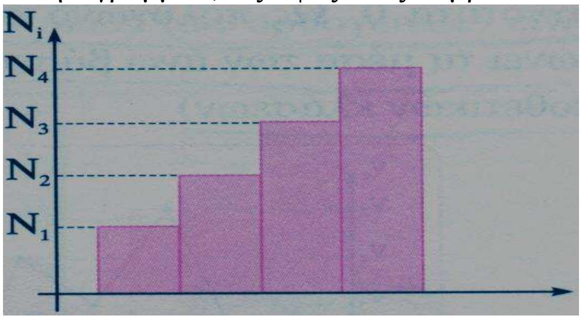


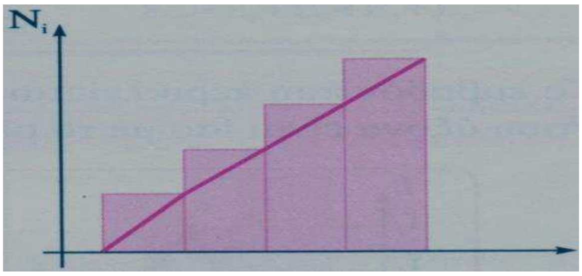


Όπως φαίνεται στο παραπάνω σχήμα, ισχύει ότι το εμβαδόν που περικλείεται από το πολύγωνο των σχετικών % συχνοτήτων και τον οριζόντιο άξονα ισούται με 100.

Ιστόγραμμα και πολύγωνο των αθροιστικών συχνοτήτων.

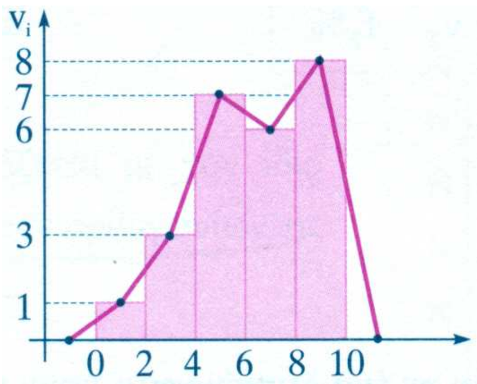
Κατά τη σχεδίαση των ορθογωνίων παραλληλογράμμων, ως ύψος τους λαμβάνονται οι αθροιστικές τους συχνότητες. Για την κατασκευή του πολυγώνου των αθροιστικών συχνοτήτων ενώνομε με ευθύγραμμα τμήματα τα δεξιά άκρα των πάνω βάσεων των ορθογωνίων παραλληλογράμμων.



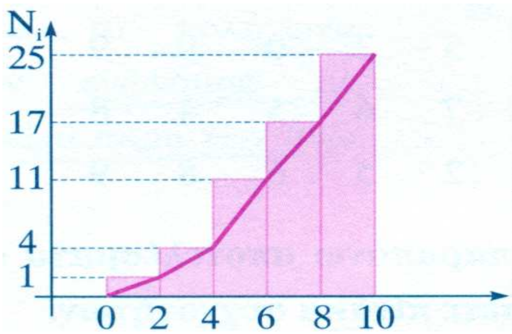


Εφαρμογή 21. Συμπληρώστε τον παρακάτω πίνακα που αναφέρεται στην ηλικία ν, 25 πλοίων. Φτιάξτε τα: ιστόγραμμα συχνοτήτων ν_i, πολύγωνο συχνοτήτων ν_i, ιστόγραμμα αθροιστικών συχνοτήτων N_i και πολύγωνο αθροιστικών συχνοτήτων N_i.

Κλάσεις	x_i	ν_i	N_i	f_i	F_i	$f_i\%$	$F_i\%$
$[0,\ 2)$	$x_1 =$	1	1				
$[2,\ 4)$	$x_2 =$	3	4				
$[4,\ 6)$	$x_3 =$	7	11				
$[6,\ 8)$	$x_4 =$	6	17				
$[8,\ 10)$	$x_5 =$	8	25				
Σύνολο		25					

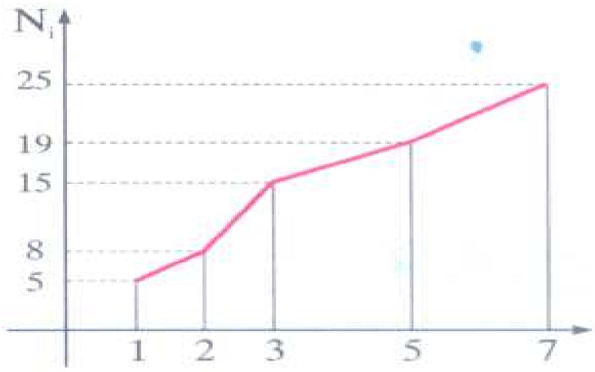


Ιστόγραμμα και πολύγωνο συχνοτήτων ν_i.



Ιστόγραμμα και πολύγωνο αθροιστικών συχνοτήτων N_i.

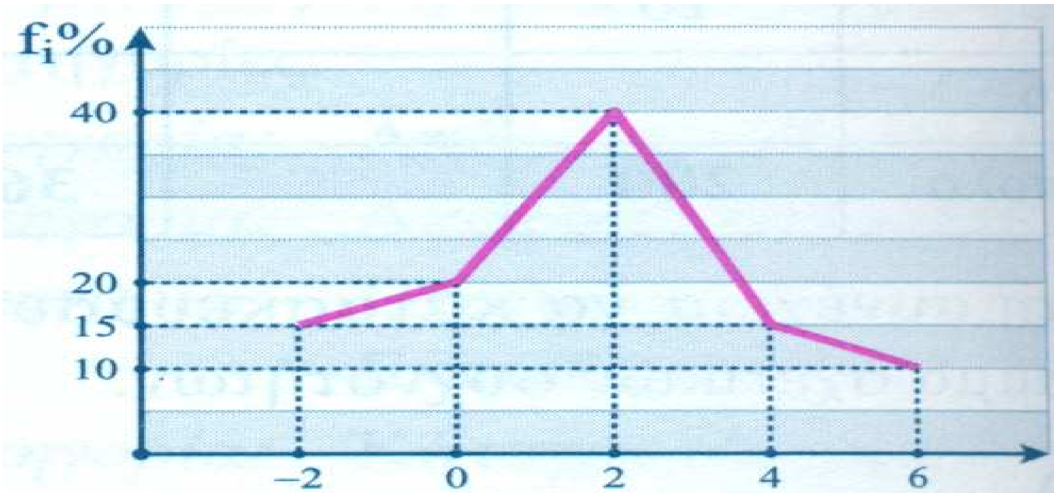
Εφαρμογή 22. Από το πολύγωνο αθροιστικών συχνοτήτων, κατασκευάστε τον πίνακα κατανομής όλων των συχνοτήτων.



Λύση.

x_i	ν_i	N_i	$f_i\%$	$F_i\%$	f_i	F_i
1	5	5	20	20	0,20	0,20
2	3	8	12	32	0,12	0,32
3	7	15	28	60	0,28	0,60
5	4	19	16	76	0,16	0,76
7	6	25	24	100	0,24	1
Σύνολο	25		100		1	

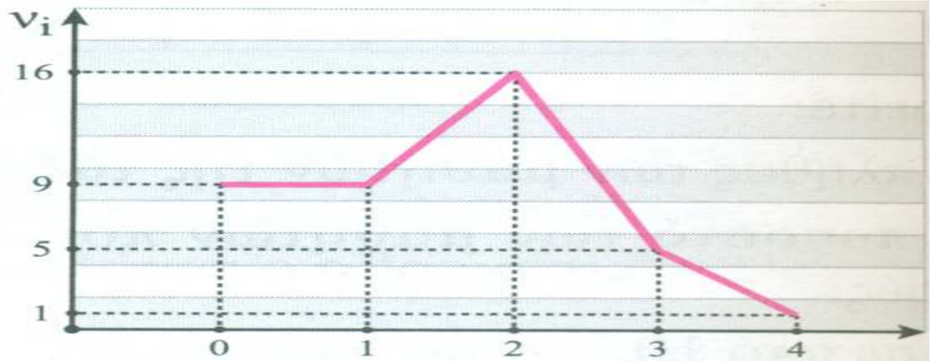
Εφαρμογή 23. Από τον πίνακα κατανομής των επί τοις εκατό σχετικών συχνοτήτων, συμπληρώστε τον πίνακα κατανομής συχνοτήτων (απολύτων, σχετικών επί τοις εκατό και αθροιστικών) για τις συνολικά 200 παρατηρήσεις της ποσοτικής μεταβλητής.



Λύση.

x_i	ν_i	$f_i\%$	N_i
−2	30	15	30
0	40	20	70
2	80	40	150
4	30	15	180
6	20	10	200
Σύνολο	200	100	

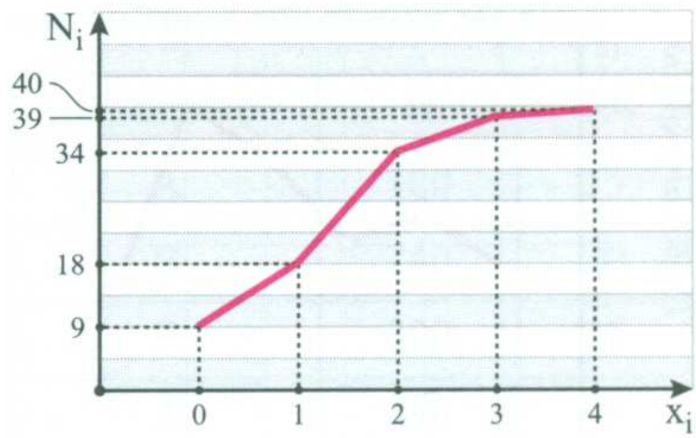
Εφαρμογή 24. Από το ακόλουθο διάγραμμα κατανομής συχνοτήτων, συμπληρώστε τον αντίστοιχο στατιστικό πίνακα και κατασκευάστε τα πολύγωνα αθροιστικών και αθροιστικών σχετικών συχνοτήτων.



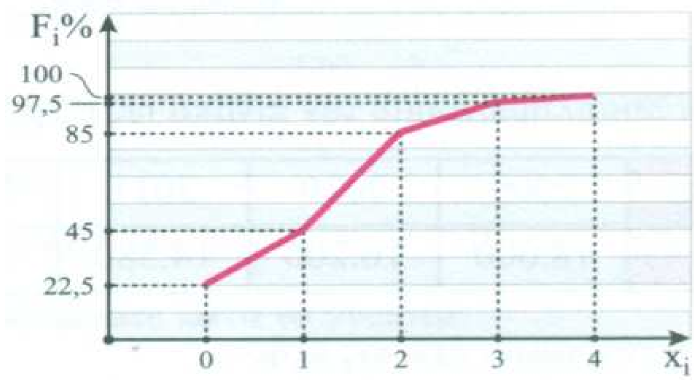
Λύση.

x_i	ν_i	N_i	$f_i\%$	$F_i\%$
0	9	9	22,5	22,5
1	9	18	22,5	45
2	16	34	40	85
3	5	39	12,5	97,5
4	1	40	2,5	100
Σύνολο	40		100	

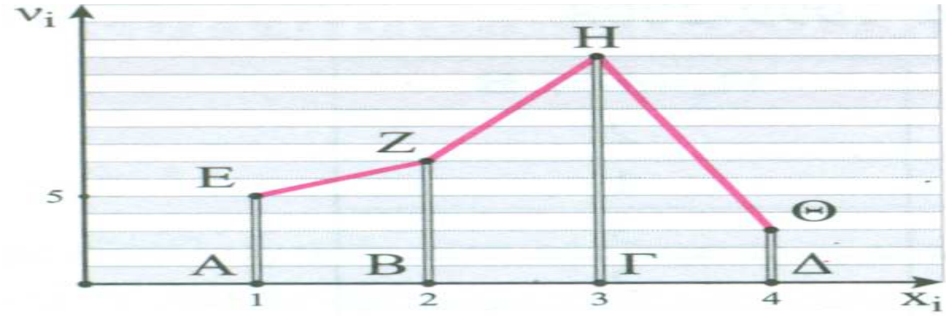
Πολύγωνο αθροιστικών συχνοτήτων.



Πολύγωνο αθροιστικών σχετικών συχνοτήτων.



Εφαρμογή 25. Από το ακόλουθο διάγραμμα συχνοτήτων, βρείτε τις απόλυτες συχνότητες ν_2, ν_3, ν_4 αν $E_{\text{ΑΒΖΕ}} = 7$, $E_{\text{ΒΓΗΖ}} = 12$, $E_{\text{ΓΔΘΗ}} = 9$.



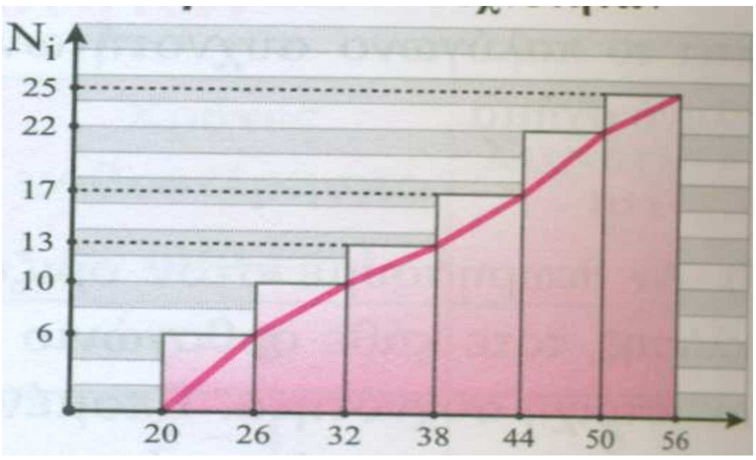
Λύση. Το εμβαδό των πολυγώνων που σχηματίζονται ισούται με το εμβαδόν των αντίστοιχων τραπεζίων. Οι βάσεις των τραπεζίων ισούται με τις αντίστοιχες συχνότητες και η βάσεις τους θεωρούνται ίσες με τη μονάδα.

$$E_{\text{ΑΒΖΕ}} = 7 \Leftrightarrow 1\frac{\nu_1+\nu_2}{2} = 7 \Leftrightarrow \frac{5+\nu_2}{2} = 7 \Leftrightarrow 5+\nu_2 = 14 \Leftrightarrow \nu_2 = 9$$

$$E_{\text{ΒΓΗΖ}} = 12 \Leftrightarrow 1\frac{\nu_2+\nu_3}{2} = 12 \Leftrightarrow \frac{9+\nu_3}{2} = 12 \Leftrightarrow 9+\nu_3 = 24 \Leftrightarrow \nu_2 = 15$$

$$E_{\text{ΓΔΘΗ}} = 9 \Leftrightarrow 1\frac{\nu_3+\nu_4}{2} = 9 \Leftrightarrow \frac{15+\nu_4}{2} = 9 \Leftrightarrow 15+\nu_4 = 18 \Leftrightarrow \nu_4 = 3$$

Εφαρμογή 26. Από τα πολύγωνο και ιστόγραμμα αθροιστικών συχνοτήτων για μία ποσοτική μεταβλητή (πραγματική θαλάσσια υπηρεσία σε μήνες) κατασκευάστε τον πίνακα συχνοτήτων.



Λύση.

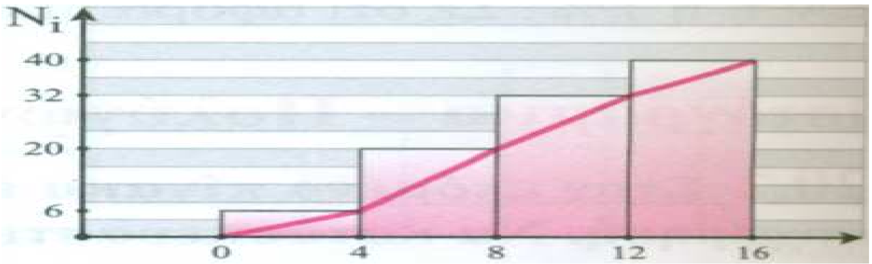
Μήνες [...−...)	N_i	ν_i	$f_i\%$
20 – 26	6	6	24
26 – 32	10	4	16
32 – 38	13	3	12
38 – 44	17	4	16
44 – 50	22	5	20
50 – 56	25	3	12
Σύνολο		25	100

Εφαρμογή 27. Ο πίνακας παρουσιάζει την ηλικία 40 πλοίων. Κατασκευάστε το πολύγωνο αθροιστικών συχνοτήτων. Πόσα πλοία έχουν ηλικία μικρότερη των:

(i) δυο,
(ii) έξι,
(iii) δεκατριών ετών;

Ηλικία πλοίων $[\ldots-\ldots)$	ν_i	N_i
0 – 4	6	6
4 – 8	14	20
8 – 12	12	32
12 – 16	8	40
Σύνολο	40	

Λύση.



Από το ιστόγραμμα αθροιστικών συχνοτήτων προκύπτει το πολύγωνο αθροιστικών συχνοτήτων.

(i) Μικρότερη ηλικία των δυο ετών, έχουν τρία πλοία $\left(\frac{\nu_1}{2}=3\right)$.

(ii) Μικρότερη ηλικία των έξι ετών, έχουν δεκατρία πλοία $\left(\nu_1+\frac{\nu_2}{2}=6+\frac{14}{2}=6+7=13\right)$.

(iii) Μικρότερη ηλικία των δεκατριών ετών έχουν τα 6 πλοία της 1ης κλάσεως, συν τα 14 πλοία της δεύτερης κλάσεως, συν τα 12 πλοία της 3ης κλάσεως, συν τα x πλοία της 4ης κλάσεως.

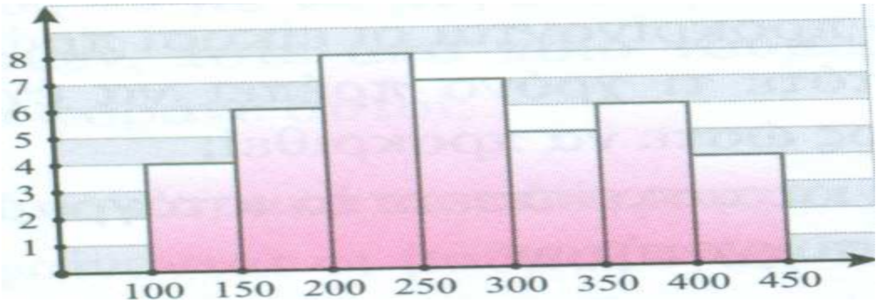
Υπολογισμός του x

Σε διάστημα πλάτους 16 – 12 αντιστοιχούν 8 πλοία.
Σε διάστημα πλάτους 13 – 12 αντιστοιχούν x πλοία.

Σε διάστημα πλάτους 4 αντιστοιχούν 8 πλοία.
Σε διάστημα πλάτους 1 αντιστοιχούν x πλοία. Άρα $x=2$.
Συνεπώς κάτω των 13 ετών είναι 6 + 14 + 12 + 2 = 34 πλοία.

2ος τρόπος υπολογισμού. Από το πολύγωνο των αθροιστικών συχνοτήτων προκύπτει ότι $\frac{40-y}{y-32}=\frac{16-13}{13-12}\Leftrightarrow\frac{40-y}{y-32}=3\Leftrightarrow 40-y=3(y-32)\Leftrightarrow y=34$.

Εφαρμογή 28. Από το ιστόγραμμα συχνοτήτων που περιγράφει τη διάρκεια ζωής ηλεκτρικών λαμπτήρων (σε ώρες), κατασκευάστε πίνακα συχνοτήτων και ιστόγραμμα αθροιστικών σχετικών % συχνοτήτων.

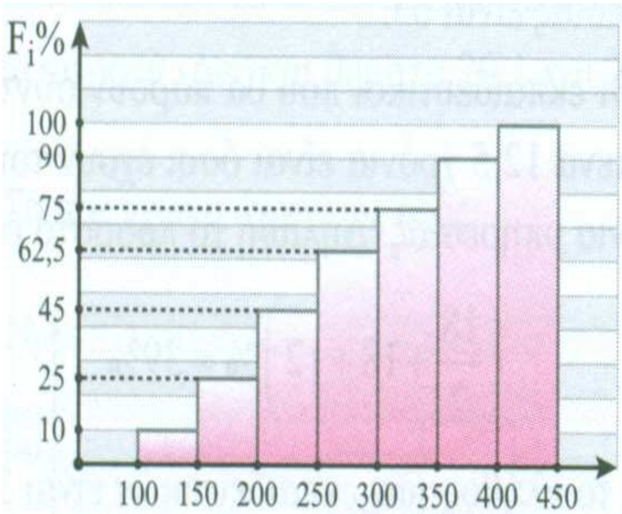


Ποιο το ποσοστό των λαμπτήρων με διάρκεια ζωής μεγαλύτερη των 320 ωρών;

Λύση.

Ώρες $[\ldots-\ldots)$	v_i	$f_i\%$	f_i	N_i	$F_i\%$	F_i
100 – 150	4	10	0,100	4	10	0,100
150 – 200	6	15	0,150	10	25	0,250
200 – 250	8	20	0,200	18	45	0,450
250 – 300	7	17,5	0,175	25	62,5	0,625
300 – 350	5	12,5	0,125	30	75	0,750
350 – 400	6	15	0,150	36	90	0,900
400 – 450	4	10	0,100	40	100	1
Σύνολο	40	100	1			

Ιστόγραμμα αθροιστικών σχετικών % συχνοτήτων.



Πάνω από 320 ώρες ζωής έχουν οι 4 λαμπτήρες της κλάσεως $[400-450)$, συν τους 6 λαμπτήρες της κλάσεως $[350-400)$ συν τους x λαμπτήρες της κλάσεως $[300-350)$

Υπολογισμός x

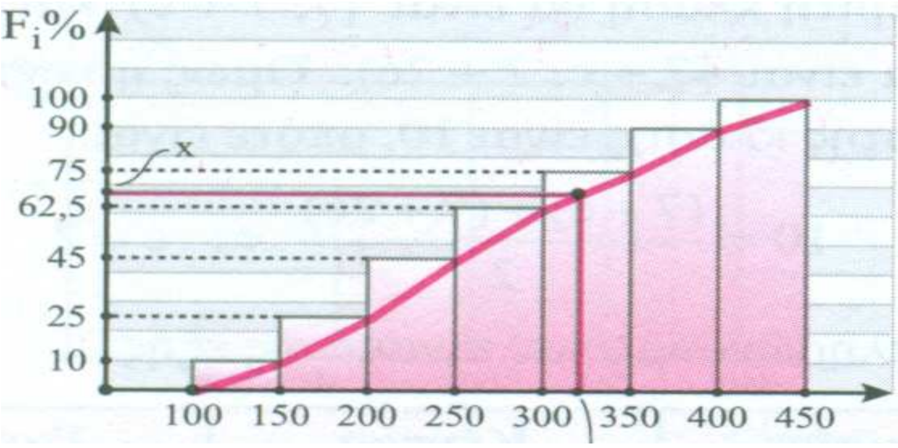
Σε διάστημα πλάτους $350-300$ αντιστοιχούν 5 λαμπτήρες.
Σε διάστημα πλάτους $320-300$ αντιστοιχούν y λαμπτήρες.

Σε διάστημα πλάτους 50 αντιστοιχούν 5 λαμπτήρες.
Σε διάστημα πλάτους 20 αντιστοιχούν y λαμπτήρες.

Άρα, $y=2$ συνεπώς, $x=3$

Οπότε 13 λαμπτήρες έχουν διάρκεια ζωής μεγαλύτερη των 320 ωρών. Αυτοί αντιστοιχούν σε ποσοστό $f_i\% = \frac{\nu_i}{\nu}100 = \frac{13}{40}100 = \frac{13}{4}10 = \frac{13}{4} = \frac{65}{2} = 32,5$.

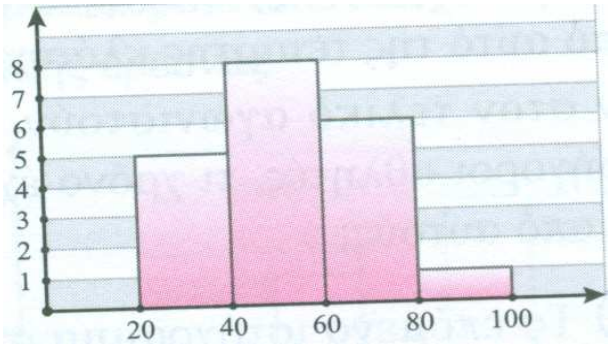
2^{ος} τρόπος υπολογισμού του ποσοστού.



$$\frac{350-320}{320-300} = \frac{75-x}{x-62,5} \Leftrightarrow \frac{30}{20} = \frac{75-x}{x-62,5} \Leftrightarrow \frac{3}{2} = \frac{75-x}{x-62,5} \Leftrightarrow x = 67,5$$

Συνεπώς, το 100–67,5= 32,5% των λαμπτήρων έχει διάρκεια ζωής μεγαλύτερη των 320 ωρών.

Εφαρμογή 29. Στο παρακάτω σχήμα φαίνονται οι βαθμοί μαθητών. Βρείτε το πλήθος των μαθητών. Πόσοι εξ΄ αυτών έλαβαν βαθμό άνω του 60;



Λύση.

Βαθμοί [...−...)	ν_i	N_i
20 – 40	5	5
40 – 60	8	13
60 – 80	6	19
80 – 100	1	20
Σύνολο	20	

Πάνω από 60 έλαβαν 6 + 1 = 7 μαθητές.

Εφαρμογή 30. Τα 100 πλοία ναυτιλιακής εταιρείας, έχουν ομαδοποιηθεί σε κλάσεις ίσου πλάτους. Έστω ότι στην κλάση $[12,\ 18)$ ανήκει το 30% των πλοίων.

Α. Ποια η απόλυτη συχνότητα των πλοίων που ανήκουν στην ανωτέρω κλάση;
Β. Πόσα πλοία της κλάσεως έχουν ηλικία από 17 έως 18 έτη;
Γ. Πόσα πλοία της κλάσεως έχουν ηλικία μικρότερη των 16 ετών;

Δ. Ποιο το ποσοστό των πλοίων της κλάσεως με ηλικίες μεταξύ 13 και 17 ετών;
Ε. Βρείτε το $x \in \mathbb{R}$ ώστε στο διάστημα $[13,\ x)$ να ανήκουν 25 πλοία.

Στ. Βρείτε το $x \in \mathbb{R}$ ώστε στο διάστημα $[13,\ x)$ να ανήκει το 15% όλων των πλοίων.

Λύση.

Α. Το 30% των συνολικά ν=100 πλοίων είναι τα 30 πλοία της κλάσεως $[12,\ 18)$.

Β. Σε διάστημα πλάτους $18-12$ αντιστοιχούν 30 πλοία
Σε διάστημα πλάτους $18-17$ αντιστοιχούν x πλοία

Συνεπώς Σε διάστημα πλάτους 6 αντιστοιχούν 30 πλοία
Σε διάστημα πλάτους 1 αντιστοιχούν x πλοία

Άρα $x=5$ πλοία.

Γ. Σε διάστημα πλάτους $18-12$ αντιστοιχούν 30 πλοία
Σε διάστημα πλάτους $16-12$ αντιστοιχούν x πλοία

Συνεπώς Σε διάστημα πλάτους 6 αντιστοιχούν 30 πλοία
Σε διάστημα πλάτους 4 αντιστοιχούν x πλοία

Άρα $x = 30\frac{4}{6} = 20$ πλοία.

Δ. Σε διάστημα πλάτους $18-12$ αντιστοιχούν 30 πλοία
Σε διάστημα πλάτους $17-13$ αντιστοιχούν x πλοία

Συνεπώς Σε διάστημα πλάτους 6 αντιστοιχούν 30 πλοία
Σε διάστημα πλάτους 4 αντιστοιχούν x πλοία

Άρα $x = 30\frac{4}{6} = 20$ πλοία. Τα $\frac{2}{3}$ των πλοίων της κλάσεως βρίσκονται μεταξύ $13-17$.

Ε. Σε διάστημα πλάτους $18-12$ αντιστοιχούν 30 πλοία
Σε διάστημα πλάτους $x-13$ αντιστοιχούν 25 πλοία
Συνεπώς Σε διάστημα πλάτους 6 αντιστοιχούν 30 πλοία
Σε διάστημα πλάτους $x-13$ αντιστοιχούν 25 πλοία
Άρα, $6 \cdot 25 = 30(x-13) \Leftrightarrow 25 = 5(x-13) \Leftrightarrow 5 = x-13 \Leftrightarrow x = 18$.

Στ. Σε διάστημα πλάτους $18-12$ αντιστοιχούν 30 πλοία
Σε διάστημα πλάτους $x-13$ αντιστοιχούν 15 πλοία
Συνεπώς Σε διάστημα πλάτους 6 αντιστοιχούν 30 πλοία
Σε διάστημα πλάτους $x-13$ αντιστοιχούν 15 πλοία
Άρα $x-13=3 \Leftrightarrow x=16$.

Εφαρμογή 31. Οι ηλικίες Δ/Ξ δίνονται στον παρακάτω στατιστικό πίνακα. Αφού συμπληρωθεί ο πίνακας, βρείτε το πλήθος και ποσοστό των πλοίων με ηλικία:

Α. κάτω των 12 ετών,

Β. άνω των $17,5$ ετών,

Γ. Τα τρία νεότερα πλοία του στόλου, είναι το πολύ έως πόσων ετών;

Δ. Που κυμαίνεται η ηλικία του γηραιότερου πλοίου του στόλου;

Ε. Ποια η μέση ηλικία όλων αυτών των δεξαμενοπλοίων;

Κλάσεις	ν_i	N_i	$f_i\%$	$F_i\%$	f_i	F_i	Κεντρικοί όροι x_i	$x_i\nu_i$
$[0,\ 5)$	$\nu_1 = 6$							
$[5,\ 10)$	$\nu_2 = 3$							
$[10,\ 15)$	$\nu_3 = 10$							
$[15,\ 20)$	$\nu_4 = 4$							
$[20,\ 25)$	$\nu_5 = 2$							
Σύνολο	$\nu = \sum_{i=1}^{5} \nu_i =$		$\sum_{i=1}^{5} f_i\% =$		$\sum_{i=1}^{5} f_i =$			$\sum_{i=1}^{5} x_i\nu_i =$

Λύση.

Κλάσεις	ν_i	N_i	$f_i\%$	$F_i\%$	f_i	F_i	Κεντρικοί όροι x_i	$x_i\nu_i$
$[0,\ 5)$	$\nu_1 = 6$	$N_1 = 6$	$f_1\% = 24$	$F_1\% = 24$	$f_1 = 0,24$	$F_1 = 0,24$	$x_1 = 2,5$	$x_1\nu_1 = 15$
$[5,\ 10)$	$\nu_2 = 3$	$N_2 = 9$	$f_2\% = 12$	$F_2\% = 36$	$f_2 = 0,12$	$F_2 = 0,36$	$x_2 = 7,5$	$x_2\nu_2 = 22,5$
$[10,\ 15)$	$\nu_3 = 10$	$N_3 = 19$	$f_3\% = 40$	$F_3\% = 76$	$f_3 = 0,40$	$F_3 = 0,76$	$x_3 = 12,5$	$x_3\nu_3 = 125$
$[15,\ 20)$	$\nu_4 = 4$	$N_4 = 23$	$f_4\% = 16$	$F_4\% = 92$	$f_4 = 0,16$	$F_4 = 0,92$	$x_4 = 17,5$	$x_4\nu_4 = 70$
$[20,\ 25)$	$\nu_5 = 2$	$N_5 = 25$	$f_5\% = 8$	$F_5\% = 100$	$f_5 = 0,08$	$F_5 = 1$	$x_5 = 22,5$	$x_5\nu_5 = 45$
Σύνολο	$\nu = \sum_{i=1}^{5}\nu_i = 25$		$\sum_{i=1}^{5} f_i\% = 100$		$\sum_{i=1}^{5} f_i = 1$			$\sum_{i=1}^{5} x_i\nu_i = 277,5$

Α. Κάτω των 12 ετών βρίσκονται τα $\nu_1 = 6$ πλοία της κλάσεως $[0,\ 5)$ συν τα $\nu_2 = 3$ πλοία της κλάσεως $[5,\ 10)$ συν τα x πλοία της κλάσεως $[10,\ 15)$ που θα υπολογίσομε.

Σε διάστημα πλάτους 15 – 10 αντιστοιχούν 10 πλοία
Σε διάστημα πλάτους 12 – 10 αντιστοιχούν x πλοία

Συνεπώς

Σε διάστημα πλάτους 5 αντιστοιχούν 10 πλοία
Σε διάστημα πλάτους 2 αντιστοιχούν x πλοία

Άρα $x = 4$ πλοία. Συνεπώς 6 + 3 + 4 = 13 πλοία.

Β. Άνω των 17,5 ετών βρίσκονται τα $\frac{\nu_4}{2} = 2$ πλοία της κλάσεως $[15,\ 20)$ συν τα $\nu_5 = 2$ πλοία της κλάσεως $[20,\ 25)$.

Γ. Τα 6 νεότερα πλοία του στόλου απαρτίζουν την κλάση $[0,\ 5)$. Συνεπώς τα 3 πιο νέα πλοία του στόλου έχουν ηλικία έως 2,5 ετών.

Δ. Τα 2 παλαιότερα του στόλου απαρτίζουν την κλάση $[20,\ 25)$. Συνεπώς το γηραιότερο πλοίο του στόλου έχει ηλικία από 22,5 έως 25 έτη.

Ε. Η μέση ηλικία όλων (και των 25) των πλοίων είναι $\bar{x} = \frac{\sum_{i=1}^{5} x_i\nu_i}{\nu} = \frac{277,5}{25} = \frac{55,5}{5} = 11,1$ έτη.

Απόλυτη συχνότητα ή συχνότητα ν_i, κατανομή συχνοτήτων, σχετική συχνότητα f_i, σχετική επί τοις εκατό συχνότητα $f_i\%$

Χαρακτηρισμός ενδεικτικού	ΑΕΝ Α΄			ΑΕΝ Β΄		
x_i	ν_i	$f_i\%$	f_i	ν_i	$f_i\%$	f_i
$x_1 =$ Μέτρια	$\nu_1 = 8$	$f_1\% = 20\%$	$f_1 = 0,2$	$\nu_1 = 6$	$f_1\% = 24\%$	$f_1 = 0,24$
$x_2 =$ Καλά	$\nu_2 = 17$	$f_2\% = 42,5\%$	$f_2 = 0,425$	$\nu_2 = 11$	$f_2\% = 44\%$	$f_2 = 0,44$
$x_3 =$ Πολύ καλά	$\nu_3 = 13$	$f_3\% = 32,5\%$	$f_3 = 0,325$	$\nu_3 = 7$	$f_3\% = 28\%$	$f_3 = 0,28$
$x_4 =$ Άριστα	$\nu_4 = 2$	$f_4\% = 5\%$	$f_4 = 0,05$	$\nu_4 = 1$	$f_4\% = 4\%$	$f_4 = 0,04$
Σύνολο	$\nu = 40$	100%	1	$\nu = 25$	100%	1

$$0 \le f_i \le 1, \qquad f_1 + f_2 + \dots + f_k = 1, \qquad f_1\% + f_2\% + \dots + f_k\% = 100\%, \qquad f_i = \frac{\nu_i}{\nu}$$

Αθροιστική συχνότητα, Σχετική αθροιστική συχνότητα

Αριθμός παιδιών που έχουν 32 οικογένειες μίας συνοικίας των Αθηνών.

x_i	ν_i	Αθροιστική συχνότητα N_i	Σχετική συχνότητα f_i	Σχετική Αθροιστική συχνότητα F_i	Σχετική συχνότητα $f_i\%$	Σχετική Αθροιστική συχνότητα F_i %
$x_1 = 0$	$\nu_1 = 4$	$N_1 = 4$	$f_1 = 0,125$	$F_1 = 0,125$	$f_1\% = 12,5\%$	$F_1\% = 12,5\%$
$x_2 = 1$	$\nu_2 = 6$	$N_2 = 10$	$f_2 = 0,1875$	$F_2 = 0,3125$	$f_2\% = 18,75\%$	$F_2\% = 31,25\%$
$x_3 = 2$	$\nu_3 = 10$	$N_3 = 20$	$f_3 = 0,3125$	$F_3 = 0,625$	$f_3\% = 31,25\%$	$F_3\% = 62,5\%$
$x_4 = 3$	$\nu_4 = 6$	$N_4 = 26$	$f_4 = 0,1875$	$F_4 = 0,8125$	$f_4\% = 18,75\%$	$F_4\% = 81,25\%$
$x_5 = 4$	$\nu_5 = 4$	$N_5 = 30$	$f_5 = 0,125$	$F_5 = 0,9375$	$f_5\% = 12,5\%$	$F_5\% = 93,75\%$
$x_6 = 5$	$\nu_6 = 2$	$N_6 = 32$	$f_6 = 2/32$	$F_6 = 1$	$f_6\% = 6,25\%$	$F_6\% = 100\%$
	$\nu = 32$		1		100%	

Εφαρμογή 32. Να συμπληρωθεί ο παρακάτω πίνακας που δείχνει το πλήθος των απορριφθέντων σπουδαστών.

Μαθήματα x_i	ν_i	$f_i\%$
Μαθηματικά	6	….
Φυσική	…	5
Ναυτιλία	8	…
ΝΗΟ	8	…
Ραντάρ	10	25
Ν. Μετεωρολογία	…	…
	…	…

Μαθήματα x_i	ν_i	$f_i\%$
Μαθηματικά	6	**15**
Φυσική	**2**	5
Ναυτιλία	8	**20**
ΝΗΟ	8	**20**
Ραντάρ	10	25
Ν. Μετεωρολογία	**6**	**15**
	1	**100**

Παράδειγμα.

Τέρματα x_i	Πλήθος ομάδων ν_i	$\nu_i x_i$
$x_1 = 0$	$\nu_1 =$	$\nu_1 x_1 = 0$
$x_2 = 1$	$\nu_2 =$	$\nu_2 x_2 =$
$x_3 = 2$	$\nu_3 =$	$\nu_3 x_3 =$
$x_4 = 3$	$\nu_4 =$	$\nu_4 x_4 =$
$x_5 = 4$	$\nu_5 =$	$\nu_5 x_5 =$
$x_6 = 5$	$\nu_6 =$	$\nu_6 x_6 =$

Ο μέσος αριθμός τερμάτων ανά ομάδα είναι

$$\bar{x} = \frac{\nu_1 x_1 + \nu_2 x_2 + \nu_3 x_3 + \nu_4 x_4 + \nu_5 x_5 + \nu_6 x_6}{\nu_1 + \nu_2 + \nu_3 + \nu_4 + \nu_5 + \nu_6} = \frac{\sum_{i=1}^{6} \nu_i x_i}{\nu_1 + \nu_2 + \nu_3 + \nu_4 + \nu_5 + \nu_6}$$

Παράδειγμα.

Εισόδημα (σε χιλιάδες €)	Πλήθος οικογενειών
$x_1 = 3$	$\nu_1 = 1$
$x_2 = 17$	$\nu_2 = 1$

Εισόδημα (σε χιλιάδες €)	Πλήθος οικογενειών
$x_1 = 9$	$\nu_1 = 1$
$x_2 = 11$	$\nu_2 = 1$

Και στις δυο περιπτώσεις είναι $\bar{x} = 10$.

$$\frac{(3-10)+(17-10)}{2} = \frac{-7+7}{2} = 0$$

Μέση απόλυτη απόκλιση e.

$$e = \frac{|3-10|+|17-10|}{2} = \frac{7+7}{2} = 7 \qquad e = \frac{|9-10|+|11-10|}{2} = \frac{1+1}{2} = 1$$

Διακύμανση ή Διασπορά

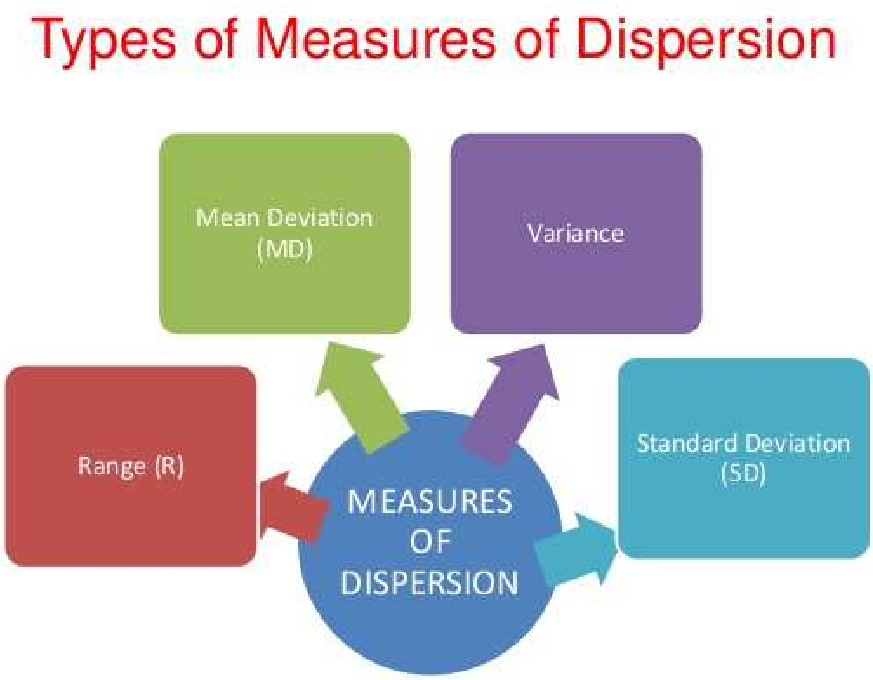
$$s^2 = \frac{(3-10)^2+(17-10)^2}{2} = \frac{49+49}{2} = 49$$

$$s^2 = \frac{(9-10)^2+(11-10)^2}{2} = \frac{1+1}{2} = 1$$

$$s^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_\nu-\bar{x})^2}{\nu} = \frac{\sum_{i=1}^{\nu}(x_i-\bar{x})^2}{\nu}$$

Για ομαδοποιημένα δεδομένα είναι

$$s^2 = \frac{\nu_1\left(x_1-\overline{x}\right)^2+\nu_2\left(x_2-\overline{x}\right)^2+\ldots+\nu_k\left(x_\nu-\overline{x}\right)^2}{\nu_1+\nu_2+\ldots+\nu_k} = \frac{\sum_{i=1}^{k}\left(x_i-\overline{x}\right)^2}{\nu}$$



Διακύμανση ή διασπορά (Variance).

Διακύμανση ή διασπορά (S^2) των παρατηρήσεων $x_1, x_2, x_3, ..., x_\nu$ ονομάζεται ο μέσος όρος των τετραγώνων των αποκλίσεων x_i από τη μέση τιμή $\overline{x}$. Δηλαδή $S^2 = \frac{1}{\nu} \cdot \sum_{i=1}^{\nu} \left(x_i - \overline{x} \right)^2$. Είναι το πιο σημαντικό μέτρο διασποράς.

• Αν έχομε λίγες παρατηρήσεις και η μέση τιμή είναι ακέραιος, χρησιμοποιούμε τον τύπο: $S^2 = \frac{1}{\nu} \cdot \sum_{i=1}^{\nu} \left(x_i - \overline{x} \right)^2$.

Εφαρμογή 33. Βρείτε τη διασπορά των παρατηρήσεων: 2, 5, 8, 11, 14.

$$\overline{x} = \frac{2+5+8+11+14}{5} = 8$$

$$S^2 = \frac{1}{5}\left[(2-8)^2 + (5-8)^2 + (8-8)^2 + (11-8)^2 + (14-8)^2 \right] =$$

$$\frac{1}{5}\left[(-6)^2 + (-3)^2 + (0)^2 + (3)^2 + (6)^2 \right] = 18$$

• Αν έχομε λίγες παρατηρήσεις και η μέση τιμή δεν είναι ακέραιος, χρησιμοποιούμε τον τύπο: $S^2 = \frac{1}{\nu}\left[\sum_{i=1}^{\nu} x_i^2 - \frac{\left(\sum_{i=1}^{\nu} x_i \right)^2}{\nu} \right] = \frac{\sum_{i=1}^{\nu} x_i^2}{\nu} - \frac{\left(\sum_{i=1}^{\nu} x_i \right)^2}{\nu^2} = \overline{x^2} - \left(\overline{x} \right)^2$.

Απόδειξη

$$s^2 = \frac{1}{\nu}\sum \left(x_i - \overline{x} \right)^2 = \frac{1}{\nu}\sum \left(x_i^2 - 2x_i\overline{x} + \overline{x}^2 \right) = \frac{1}{\nu}\left(\sum x_i^2 - \sum 2x_i\overline{x} + \sum \overline{x}^2 \right)$$

$$= \frac{1}{\nu}\left(\sum x_i^2 - 2\overline{x}\sum x_i + \nu\overline{x}^2 \right) = \frac{1}{\nu}\left(\sum x_i^2 - 2\frac{\sum x_i}{\nu}\sum x_i + \nu\frac{\sum x_i}{\nu}\frac{\sum x_i}{\nu} \right)$$

$$= \frac{1}{\nu}\left(\sum x_i^2 - 2\frac{\left(\sum x_i \right)^2}{\nu} + \frac{\left(\sum x_i \right)^2}{\nu} \right) = \frac{1}{\nu}\left(\sum x_i^2 - \frac{\left(\sum x_i \right)^2}{\nu} \right)$$

Εφαρμογή 34. Βρείτε τη διασπορά των παρατηρήσεων $t_1 = 1,\ t_2 = 3,\ t_3 = 4,\ t_4 = 5,\ t_5 = 6$.

Είναι $\sum_{i=1}^{\nu} t_i = 1+3+4+5+6 = 19$ & $t_1^2 = 1,\ t_2^2 = 9,\ t_3^2 = 16,\ t_4^2 = 25,\ t_5^2 = 36$. Άρα $\sum_{i=1}^{\nu} t_i^2 = 1+9+16+25+36 = 87$. Άρα $S^2 = \frac{1}{\nu}\left[\sum_{i=1}^{\nu} t_i^2 - \frac{\left(\sum_{i=1}^{\nu} t_i \right)^2}{\nu} \right] = \frac{1}{5}\left(87 - \frac{19^2}{5} \right) = 2,96$.

Δηλαδή είναι $s^2 = \overline{x^2} - \left(\bar{x}\right)^2 = \frac{87}{5} - \left(\frac{19}{5}\right)^2 = 2,96$.

- Αν έχομε πίνακα συχνοτήτων ή ομαδοποιημένα δεδομένα χρησιμοποιούμε τον τύπο: $S^2 = \frac{1}{\nu}\left[\sum_{i=1}^{\kappa} x_i^2\nu_i - \frac{\left(\sum_{i=1}^{\kappa} x_i\nu_i\right)^2}{\nu}\right] = \frac{\sum_{i=1}^{\kappa} x_i^2\nu_i}{\nu} - \frac{\left(\sum_{i=1}^{\kappa} x_i\nu_i\right)^2}{\nu^2} = \frac{\sum_{i=1}^{\kappa} x_i^2\nu_i}{\nu} - \left(\frac{\sum_{i=1}^{\kappa} x_i\nu_i}{\nu}\right)^2$.

Στην περίπτωση των ομαδοποιημένων δεδομένων τη θέση των x_i παίρνουν τα κέντρα των κλάσεων.

Απόδειξη.

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

$$= \frac{1}{\nu}\sum_{i=1}^{k}\nu_i x_i^2 - \frac{2\bar{x}}{\nu}\sum_{i=1}^{k}\nu_i x_i + \frac{\bar{x}^2}{\nu}\sum_{i=1}^{k}\nu_i == \frac{1}{\nu}\sum_{i=1}^{k}\nu_i x_i^2 - \frac{2\bar{x}}{\cancel{\nu}}\cancel{\nu}\bar{x} + \frac{\bar{x}^2}{\cancel{\nu}}\cancel{\nu}$$

$$= \frac{1}{\nu}\sum_{i=1}^{k}\nu_i x_i^2 - 2\bar{x}^2 + \bar{x}^2 = \frac{1}{\nu}\sum_{i=1}^{k}\nu_i x_i^2 - \bar{x}^2 = \frac{1}{\nu}\sum_{i=1}^{k}x_i^2\nu_i - \frac{\left(\sum_{i=1}^{k}x_i\nu_i\right)^2}{\nu^2} = \frac{1}{\nu}\left\{\sum_{i=1}^{k}x_i^2\nu_i - \frac{\left(\sum_{i=1}^{k}x_i\nu_i\right)^2}{\nu}\right\}$$

Εφαρμογή 35. Βρείτε τη διασπορά των παρατηρήσεων για τα δεδομένα του παρακάτω στατιστικού πίνακα:

x_i	ν_i
2	3
3	2
5	4
6	1
Σύνολο	10

Για πρακτικούς λόγους κατασκευάζομε τον παρακάτω πίνακα.

x_i	ν_i	x_i^2	$x_i^2\nu_i$	$x_i\nu_i$
2	3	4	12	6
3	2	9	18	6
5	4	25	100	20
6	1	36	36	6
Σύνολο	10		$\sum_{i=1}^{\kappa} x_i^2\nu_i = 166$	$\sum_{i=1}^{\kappa} x_i\nu_i = 38$

Με αντικατάσταση προκύπτει $S^2 = \frac{1}{\nu}\left[\sum_{i=1}^{\kappa} x_i^2\nu_i - \frac{\left(\sum_{i=1}^{\kappa} x_i\nu_i\right)^2}{\nu}\right] = \frac{1}{10}\left[166 - \frac{38^2}{10}\right] = 2,16$.

Δηλαδή $S^2 = \dfrac{\sum_{i=1}^{\kappa} x_i^2 \nu_i}{\nu} - \left(\dfrac{\sum_{i=1}^{\kappa} x_i \nu_i}{\nu} \right)^2 = \dfrac{166}{10} - \left(\dfrac{38}{10}\right)^2 = 2{,}16$.

Πλεονεκτήματα διασποράς.	**Μειονεκτήματα διασποράς.**
Λαμβάνονται υπόψη στον υπολογισμό τους όλες οι παρατηρήσεις.	Χρειάζονται πολύ περισσότερες αλγεβρικές πράξεις για τον υπολογισμό παρά στα άλλα μέτρα.
Έχουν μεγάλη εφαρμογή.	Η διασπορά δεν εκφράζεται στις ίδιες μονάδες με το χαρακτηριστικό, κάτι το οποίο παύει να ισχύει όταν χρησιμοποιήσομε την τυπική απόκλιση.
Σε κανονικούς πληθυσμούς το 68%, 95%, 99,7% των παρατηρήσεων βρίσκονται στα $\left(\bar{x}-s,\ \bar{x}+s\right)$, $\left(\bar{x}-2s,\ \bar{x}+2s\right)$ $\left(\bar{x}-3s,\ \bar{x}+3s\right)$ διαστήματα αντίστοιχα.	

Παρατηρήσεις.

• Έστω μεταβλητή που παίρνει τις τιμές 1, 2, 3, 4, 5. Είναι $\bar{x}=3$, $s^2=2{,}5$. Αν οι αρχικές τιμές της μεταβλητής πολλαπλασιασθούν με μία σταθερά a, έστω $a=10$, οι νέες τιμές γίνονται 10, 20, 30, 40, 50 και είναι $\bar{x}=30$, $s^2=250$.

Δηλαδή ο μέσος όρος πολλαπλασιάσθηκε επί τη σταθερά $a=10$, ενώ η διασπορά πολλαπλασιάσθηκε επί $a^2=100$.

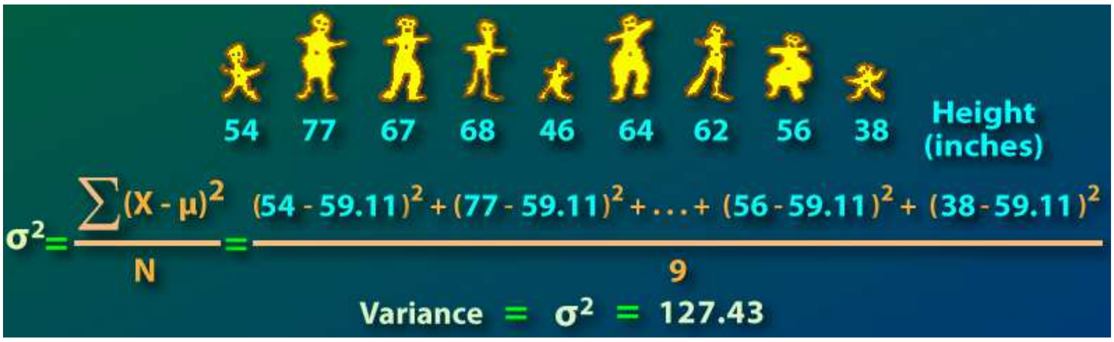
• Εξετάζομε τους δευτεροετείς σπουδαστές ως προς τον αριθμό των σελίδων μαθηματικών που μελέτησαν την προηγούμενη εβδομάδα. Έστω ότι τα μέτρα θέσεως είναι: $\bar{x}$, M_0, δ, P_{70}. Έστω ότι τα μέτρα διασποράς είναι: s, Q.

Αν όλοι οι σπουδαστές μελετήσουν επιπλέον 100 σελίδες, τα νέα μέτρα θέσεως και διασποράς είναι:

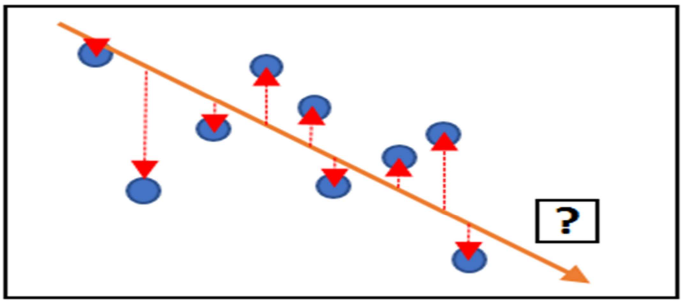
$\bar{x}' = \bar{x}+100$, $s'=s$, $\mathrm{M}=\mathrm{M}_0+100$, $\delta'=\delta+100$, $Q'=Q$, $P_{70}'=P_{70}+100$.

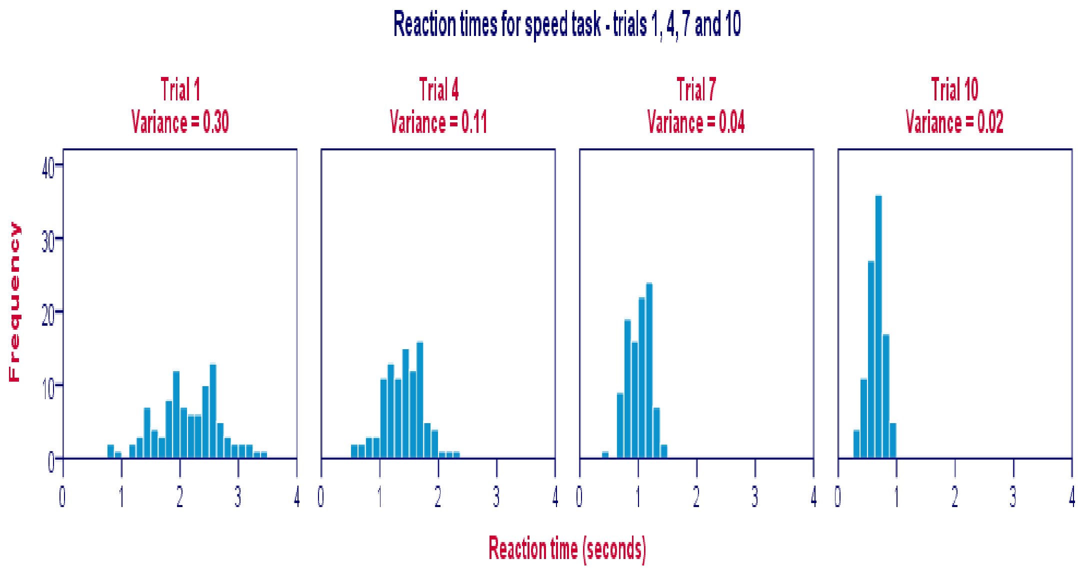
• Γενικά ισχύουν τα παρακάτω:

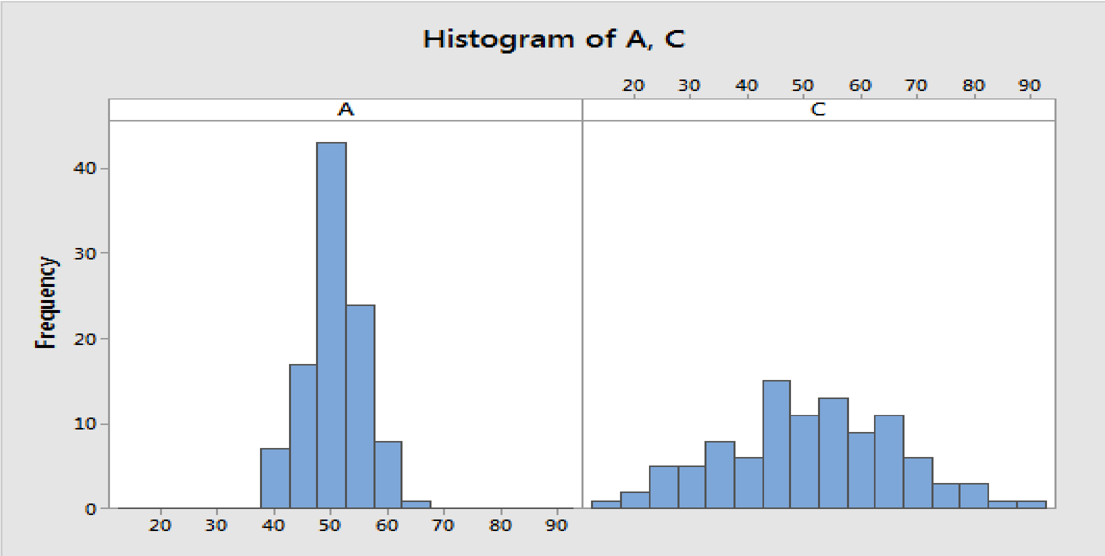
x_1	x_2	x_3	$\bar{x}$	s^2	s	M_0	δ	Q	P_{70}
x_1+c	x_2+c	x_3+c	$\bar{x}+c$	s^2	s	M_0+c	$\delta+c$	Q	$P_{70}+100$
kx_1	kx_2	kx_3	$k\bar{x}$	k^2s^2	$\lvert k\rvert s$		$k\delta$		

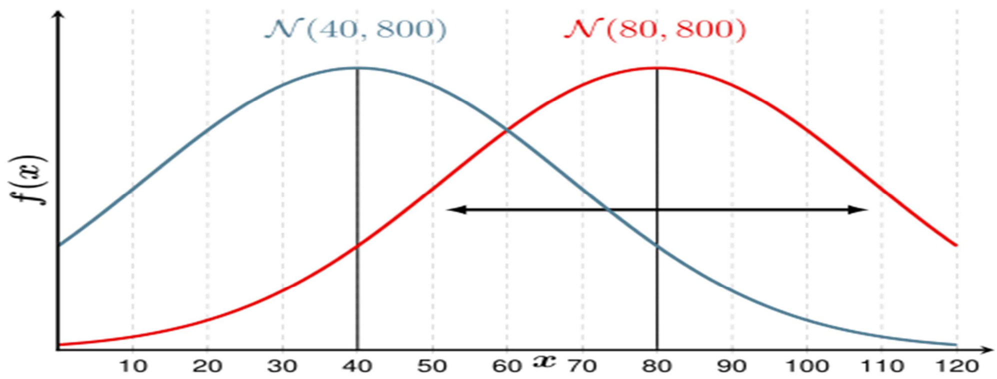

54 77 67 68 46 64 62 56 38
Height (inches)
$\sigma^2 = \frac{\sum(X-\mu)^2}{N} = \frac{(54-59.11)^2+(77-59.11)^2+\ldots+(56-59.11)^2+(38-59.11)^2}{9}$
Variance = σ2 = 127.43


54 77 67 68 46 64 62 56 38
Height (inches)
Variance = σ2 = 127.43
σ = √127.43 = 11.29″
Standard Deviation = 11.29″
μ = 59.11″
0 10 20 30 40 50 60 70 80 90
Scale of Measures (inches)


?


Reaction times for speed task - trials 1, 4, 7 and 10
Trial 1
Variance = 0.30
Trial 4
Variance = 0.11
Trial 7
Variance = 0.04
Trial 10
Variance = 0.02
Frequency
Reaction time (seconds)


Histogram of A, C
A
C
Frequency
0
10
20
30
40
20
30
40
50
60
70
80
90


$\mathcal{N}(40, 800)$
$\mathcal{N}(80, 800)$
$f(x)$
x
10
20
30
40
50
60
70
80
90
100
110
120

Εφαρμογή 36. Υπολογίστε τα μέτρα διασποράς της παρακάτω κατανομής των βαθμών που έλαβε σπουδαστής σε είκοσι μαθήματα, αφού πρώτα συμπληρώσετε τον πίνακα.

Βαθμοί Μαθημάτων $[\ldots-\ldots)$	x_i	ν_i	$x_i\nu_i$
$0-2$	$x_1=1$	$\nu_1=4$	$x_1\nu_1=$
$2-4$	$x_2=3$	$\nu_2=10$	$x_2\nu_2=$
$4-6$	$x_3=5$	$\nu_3=3$	$x_3\nu_3=$
$6-8$	$x_4=7$	$\nu_4=2$	$x_4\nu_4=$
$8-10$	$x_5=9$	$\nu_5=1$	$x_5\nu_5=$
Σύνολο		$\nu=\sum_{i=1}^{5}\nu_i=$	$\sum_{i=1}^{5}x_i\nu_i=$

Λύση.

Βαθμοί Μαθημάτων $[\ldots-\ldots)$	x_i	ν_i	$x_i\nu_i$
$0-2$	$x_1=1$	$\nu_1=4$	$x_1\nu_1=4$
$2-4$	$x_2=3$	$\nu_2=10$	$x_2\nu_2=30$
$4-6$	$x_3=5$	$\nu_3=3$	$x_3\nu_3=15$
$6-8$	$x_4=7$	$\nu_4=2$	$x_4\nu_4=14$
$8-10$	$x_5=9$	$\nu_5=1$	$x_5\nu_5=9$
Σύνολο		$\nu=\sum_{i=1}^{5}\nu_i=20$	$\sum_{i=1}^{5}x_i\nu_i=72$

$R=10-0=10$ $\qquad$ $\bar{x}=\dfrac{\sum_{i=1}^{5}x_i\nu_i}{\nu}=\dfrac{72}{20}=\dfrac{36}{10}=3,6$

Η διάμεσος κυμαίνεται μεταξύ 3 και 4.

Επικρατούσα κλάση είναι η $2-4$.

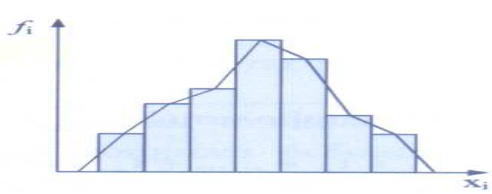
$$s^2=\frac{\sum_{i=1}^{5}\left(x_i-\bar{x}\right)^2\cdot\nu_i}{\nu}=\frac{\left(x_1-\bar{x}\right)^2\cdot\nu_1+\left(x_2-\bar{x}\right)^2\cdot\nu_2+\left(x_3-\bar{x}\right)^2\cdot\nu_3+\left(x_4-\bar{x}\right)^2\cdot\nu_4+\left(x_5-\bar{x}\right)^2\cdot\nu_5}{\nu}=$$

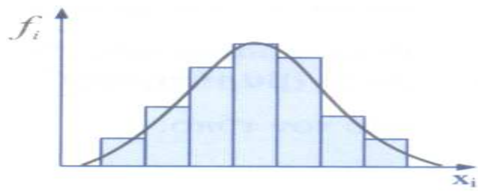
$$\frac{(1-3,6)^2\cdot4+(3-3,6)^2\cdot10+(5-3,6)^2\cdot3+(7-3,6)^2\cdot2+(9-3,6)^2\cdot1}{20}=$$

$$\frac{2,6^2\cdot4+0,36\cdot10+1,4^2\cdot3+(3,4)^2\cdot2+5,4^2}{20}=\ldots$$

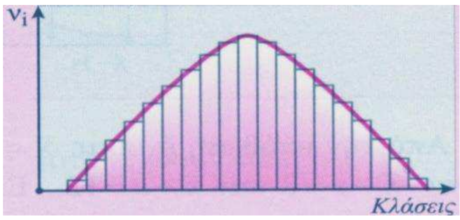
Τυπική απόκλιση (Standard deviation).

Τυπική απόκλιση (S) ονομάζεται η θετική τετραγωνική ρίζα της διακυμάνσεως ή διασποράς (variance). Δηλαδή $S=\sqrt{S^2}$. Εκφράζεται με τις ίδιες μονάδες που εκφράζονται και οι παρατηρήσεις. Έστω η συνεχής μεταβλητή X ενός πληθυσμού. Χωρίζεται το δείγμα σε μεγάλο αριθμό κλάσεων, πολύ μικρού πλάτους η κάθε μία. Παρατηρούμε ότι το πολύγωνο συχνοτήτων τείνει να γίνει ομαλή καμπύλη που ονομάζεται καμπύλη συχνοτήτων.

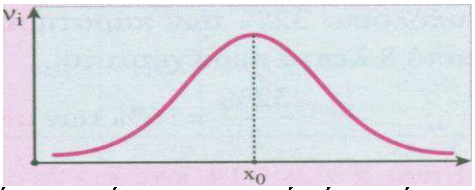




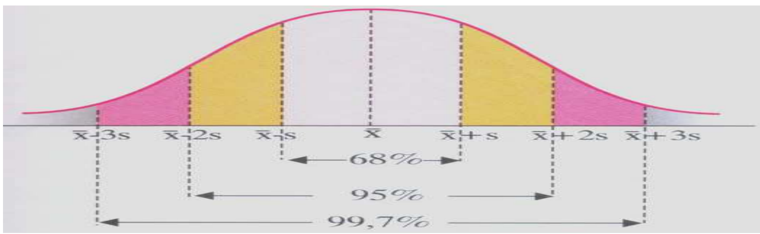
Καμπύλη συχνοτήτων (Frequency curve) είναι η μορφή που θα έπαιρνε ένα πολύγωνο συχνοτήτων, αν το πλήθος των κλάσεων ήταν πολύ μεγάλο και το πλάτος των κλάσεων πολύ μικρό.

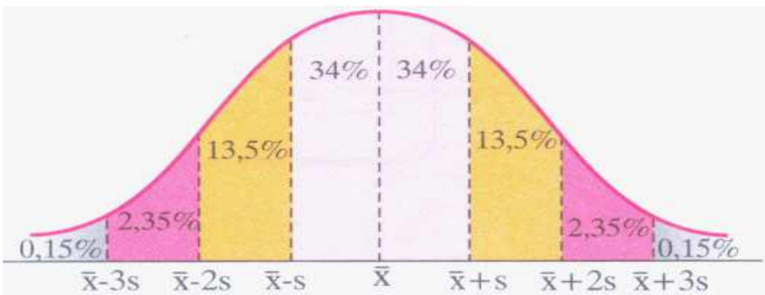


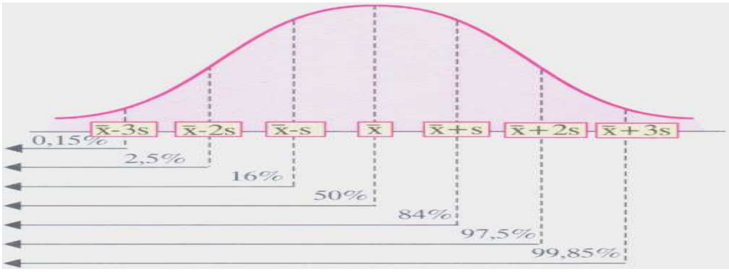
Όταν η καμπύλη συχνοτήτων λάβει τη μορφή του διπλανού σχήματος, η κατανομή συχνοτήτων ονομάζεται κανονική. Δηλαδή η μεταβλητή Χ ακολουθεί την κανονική κατανομή (Normal distribution). Αν η καμπύλη συχνοτήτων, για το χαρακτηριστικό που εξετάσαμε είναι κανονική ή περίπου κανονική, τότε η τυπική απόκλιση s έχει τις παρακάτω ιδιότητες:

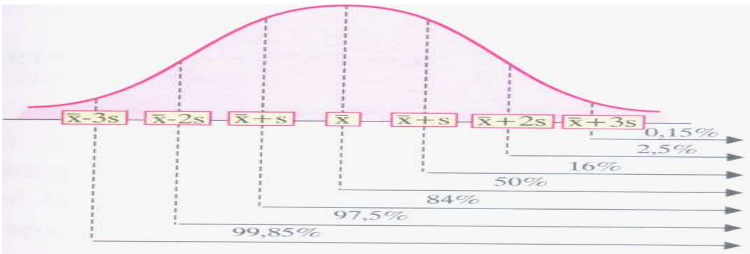


- Το 68% περίπου των παρατηρήσεων βρίσκεται στο διάστημα $\left(\bar{x}-s,\ \bar{x}+s\right)$.
- Το 95% περίπου των παρατηρήσεων βρίσκεται στο διάστημα $\left(\bar{x}-2s,\ \bar{x}+2s\right)$.
- Το 99,7% περίπου των παρατηρήσεων βρίσκεται στο διάστημα $\left(\bar{x}-3s,\ \bar{x}+3s\right)$.
- Το εύρος ισούται περίπου με έξι τυπικές αποκλίσεις, δηλαδή $R\approx 6\cdot s$.



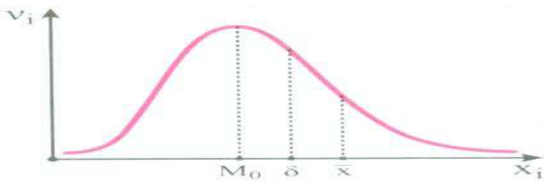




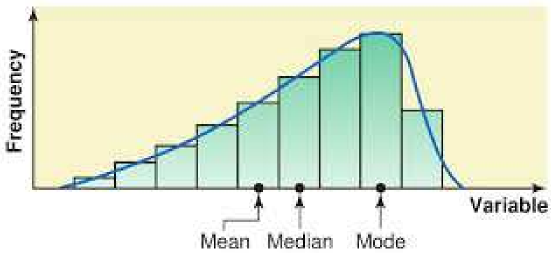


Στην κανονική κατανομή, ισχύουν $\bar{x} = \delta$ δηλαδή η μέση τιμή συμπίπτει με τη διάμεσο και $R \cong 6s$, δηλαδή το εύρος είναι περίπου εξαπλάσιο της τυπικής αποκλίσεως. Αν M_0 η επικρατούσα τιμή, τότε είναι $M_0 = \bar{x} = \delta$.

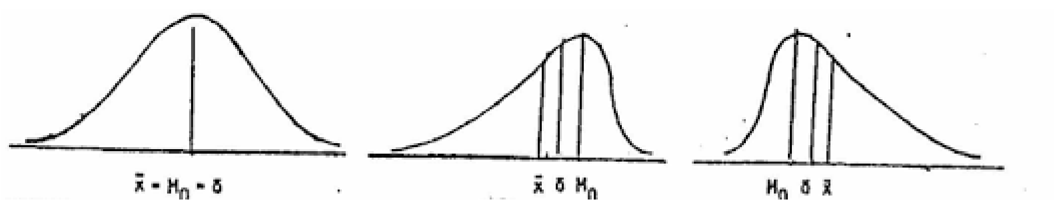
Αν οι παρατηρήσεις δεν είναι συμμετρικά κατανεμημένες, έχομε ασύμμετρη κατανομή. Αν η κατανομή παρουσιάζει θετική ασυμμετρία ισχύει ότι $M_0 < \delta < \bar{x}$.



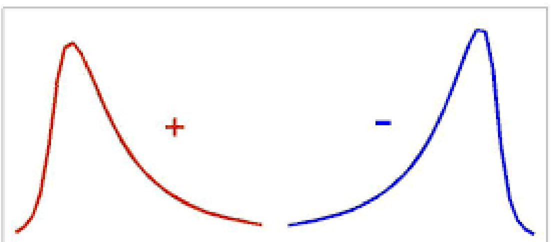
Αν η κατανομή παρουσιάζει αρνητική ασυμμετρία ισχύει ότι $\bar{x} < \delta < M_0$.



Η σχέση μέσης τιμής, διαμέσου και επικρατούσας τιμής στις διάφορες καμπύλες συχνοτήτων είναι:



Χαρακτηριστικές καμπύλες συχνοτήτων.



Συντελεστής μεταβολής ή μεταβλητότητας (coefficient of variation).

Όταν μελετάμε δυο δείγματα ως προς την ίδια μεταβλητή, παρουσιάζονται διαφορετικές τιμές στις παραμέτρους θέσεως και διασποράς.

Η μέση τιμή των υψών των ναυτικών εταιρείας Α είναι $\overline{x} = 160\ cm$ και η τυπική απόκλιση είναι $s = 5\ cm$. Οι αντίστοιχες τιμές για τους ναυτικούς της εταιρείας Β είναι $\overline{x} = 180\ cm$ και $s = 5\ cm$. Προφανώς οι δυο πληθυσμοί, δεν παρουσιάζουν τον ίδιο βαθμό μεταβλητότητας.

Υπάρχει περίπτωση να μετράμε τους δυο πληθυσμούς σε διαφορετικές μονάδες ή κλίμακες. Οι φοιτητές ελληνικού ΑΕΙ (κλίμακα βαθμολογίας 0–10) εξεταζόμενοι ως προς την επίδοση τους στη στατιστική, παρουσιάζουν μέση τιμή $\overline{x} = 7$ και τυπική απόκλιση $s = 2$. Οι φοιτητές ΑΕΙ της αλλοδαπής (κλίμακα βαθμολογίας 0–100) εξεταζόμενοι ως προς την επίδοση τους στη στατιστική, παρουσιάζουν μέση τιμή $\overline{x} = 65$ και τυπική απόκλιση $s = 20$.

Λόγω της διαφορετικής κλίμακας δεν είναι δυνατό να συγκριθούν τα ανωτέρω δυο δείγματα. Προκειμένου να ξεπερασθεί η δυσκολία αυτή, χρησιμοποιείται ο συντελεστής μεταβολής. Ο συντελεστής μεταβολής δίνεται από τον τύπο $\mathrm{CV} = \frac{s}{|\overline{x}|}$ ή αν εκφράζεται ως ποσοστό, δίνεται από τον τύπο $\mathrm{CV}\% = \frac{s}{|\overline{x}|}\%$.

Ακριβέστερα ορίζεται ως $cv = \frac{\sigma}{\mu}$ για πληθυσμό και ως $cv = \frac{s}{\overline{x}}$ για δείγμα.

Χρησιμοποιείται για τη σύγκριση της μεταβλητότητας δυο ή περισσοτέρων ομάδων τιμών (κατανομών) που εκφράζονται σε διαφορετική μονάδα μετρήσεως. Ουσιαστικά μετρά την ομοιογένεια του πληθυσμού. Αν η τιμή του είναι κάτω από το 10%, ο πληθυσμός του δείγματος θεωρείται ομοιογενής. Είναι δηλαδή μέτρο σχετικής μεταβλητότητας. Μέτρα απόλυτης μεταβλητότητας είναι τα $R,\ Q,\ \sigma,\ \sigma^2$.

Εφαρμογή 37. Η τιμή της μετοχής της εταιρείας Α το 2.000 είχε μέση τιμή 5.000 δρχ. και τυπική απόκλιση 1.000 δρχ. Οι αντίστοιχες τιμές για τη Β εταιρεία ήταν 12.000 δρχ. και 4.000 δρχ. αντίστοιχα. Ποια μετοχή είχε περισσότερη μεταβλητότητα; $cv_{\mathrm{A}} = \frac{1.000}{5.000}100 = 20\ \%$ $\quad cv_{\mathrm{B}} = \frac{4.000}{12.000}100 = 33,\overline{3}\ \%$

Πλεονεκτήματα του CV.	**Μειονεκτήματα του CV.**
Χρησιμοποιείται ως μέτρο ομοιογένειας ενός πληθυσμού.	Δεν ενδείκνυται, αν η μέση τιμή είναι κοντά στο μηδέν.
Χρησιμοποιείται ακόμα και σε διαφορετικές μονάδες μέτρησης.	
Είναι καθαρός αριθμός.	

Παρατηρήσεις.

- Ο CV αποτελεί μέτρο διασποράς. Εκφράζει την τυπική απόκλιση στο ποσοστό της μέσης τιμής $\overline{x}$.
- Ο CV εκφράζεται επί τοις εκατό (%) και είναι καθαρός αριθμός, δηλαδή ανεξάρτητος από τη μονάδα μετρήσεως.
- Ο CV είναι αξιόπιστο μέτρο μόνο αν η μέση τιμή $\overline{x}$ δε λαμβάνει τιμές κοντά στο μηδέν.

• Ομογενές ονομάζεται το δείγμα αν ο CV είναι μικρότερος ή ίσος του 10% .
• Αν ισχύει ότι $s > |\overline{x}|$ τότε ο CV λαμβάνει τιμές μεγαλύτερες της μονάδος.
• Μεταξύ δυο δειγμάτων αυτό που έχει μικρότερο συντελεστή μεταβολής έχει μεγαλύτερη ομοιογένεια.
• Μεταξύ δυο δειγμάτων που έχουν την ίδια διασπορά, αυτό που έχει μεγαλύτερη μέση τιμή $\overline{x}$, θα έχει μικρότερο συντελεστή μεταβολής.
• Όσο πιο μεγάλη διασπορά έχει ένα δείγμα, τόσο μικρότερη είναι η ομοιογένεια του και αντίστροφά.

Εφαρμογή 38. Το μέσο πλήθος των υπαλλήλων 30 ναυτιλιακών εταιρειών με έδρα τον Πειραιά που ενδεικτικά εξετάσθηκαν είναι $\overline{x_\Pi} = 50$ με $s_\Pi = 7$.

Αντιστοίχως για 20 ναυτιλιακές εταιρείες με έδρα το Λονδίνο που ενδεικτικά εξετάσθηκαν είναι $\overline{x_\Lambda} = 75$ με $s_\Lambda = 7$.

Αφού υπολογίσετε τους συντελεστές μεταβολής των δυο δειγμάτων, εξετάστε τα ως προς την ομοιογένεια τους.

Λύση.

$CV_\Pi = \dfrac{s_\Pi}{\overline{x_\Pi}} = \dfrac{7}{50} = \dfrac{14}{100} = 0,14$ $\quad CV_\Pi = 14\%$ όχι ομοιογενές το δείγμα

$CV_\Lambda = \dfrac{s_\Lambda}{\overline{x_\Lambda}} = \dfrac{7}{75} = 0,093$ $\quad CV_\Lambda = 9,3\%$ ομοιογενές το δείγμα

$CV_\Lambda < CV_\Pi$

Εφαρμογή 39. Έστω κανονική κατανομή $N(\overline{x},\ s^2) = N(30,\ 25)$ με $\overline{x} = 30$, $s = 5$. Ποια ποσοστά των παρατηρήσεων ανήκουν στα διαστήματα: $(25,\ 35)$, $(20,\ 40)$, $(15,\ 45)$; Είναι το δείγμα ομοιογενές;

Λύση.

	$\overline{x}-3s$	$\overline{x}-2s$	$\overline{x}-s$	$\overline{x}$	$\overline{x}+s$	$\overline{x}+2s$	$\overline{x}+3s$	
	15	20	25	30	35	40	45	
0,15%	2,35%	13,5%	34%	34%	13,5%	2,35%	0,15%	

Στο διάστημα $(25,\ 35) = (\overline{x}-s,\ \overline{x}+s)$ ανήκει το 68% των παρατηρήσεων.

Στο διάστημα $(20,\ 40) = (\overline{x}-2s,\ \overline{x}+2s)$ ανήκει το 95% των παρατηρήσεων.

Στο διάστημα $(15,\ 45) = (\overline{x}-3s,\ \overline{x}+3s)$ ανήκει το 99,7% των παρατηρήσεων.

Είναι $CV = \dfrac{s}{\overline{x}} = \dfrac{5}{30} = \dfrac{1}{6} > \dfrac{1}{10}$. Συνεπώς το δείγμα δεν είναι ομοιογενές.

Εφαρμογή 40. Έστω ότι οι θερμοκρασίες που αναπτύσσονται στο εσωτερικό μηχανής, κατά τη λειτουργία της, είναι πάντα μεγαλύτερες από $0^o C$ και ακολουθούν την κανονική κατανομή με μέση θερμοκρασία $100^o C$ και $CV = 10\%$.

Υπολογίστε το ποσοστό επί των ωρών λειτουργίας της μηχανής που η αναπτυσσόμενη θερμοκρασία στο εσωτερικό της είναι:

- μικρότερη των $80^o C$,
- μεγαλύτερη των $130^o C$,
- μεταξύ των $110^o C$ και $120^o C$.

Λύση.

Είναι $\left.\begin{array}{l} CV = 10\% = \dfrac{10}{100} \\ CV = \dfrac{s}{\bar{x}} = \dfrac{s}{100} \end{array}\right\} \dfrac{s}{100} = \dfrac{10}{100} \Rightarrow s = 10$

	$\bar{x}-3s$		$\bar{x}-2s$		$\bar{x}-s$		$\bar{x}$		$\bar{x}+s$		$\bar{x}+2s$		$\bar{x}+3s$	
	70		80		90		100		110		120		130	
0,15%		2,35%		13,5%		34%		34%		13,5%		2,35%		0,15%

Μικρότερη των $80^o C$ είναι η θερμοκρασία κατά το 0,15 + 2,35 = 2,5% των ωρών λειτουργίας της μηχανής. Μεγαλύτερη των $130^o C$ είναι η θερμοκρασία κατά το 0,15% των ωρών λειτουργίας της μηχανής. Μεταξύ των $110^o C$ και $120^o C$ είναι η θερμοκρασία κατά το 13,5% των ωρών λειτουργίας της μηχανής.

Εφαρμογή 41. Ο χρόνος αναμονής πολιτών έως ότου εξυπηρετηθούν σε μία δημόσια υπηρεσία, ακολουθεί την κανονική κατανομή με μέση τιμή 5 λεπτά και τυπική απόκλιση 1 λεπτό. Υπολογίστε, κατά προσέγγιση, το ποσοστό πολιτών που εξυπηρετούνται σε χρόνο:

- από 4 έως 6 λεπτά
- από 3 έως 6 λεπτά

Υπολογίστε διάμεσο και εύρος της κατανομής του χρόνου αναμονής πολιτών. Υπολογίστε συντελεστή μεταβολής της κατανομής του χρόνου αναμονής.

Λύση.

	$\bar{x}-3s$		$\bar{x}-2s$		$\bar{x}-s$		$\bar{x}$		$\bar{x}+s$		$\bar{x}+2s$		$\bar{x}+3s$	
	2		3		4		5		6		7		8	
0,15%		2,35%		13,5%		34%		34%		13,5%		2,35%		0,15%

Από 4 έως 6 λεπτά είναι ο χρόνος που χρειάζεται ώστε να εξυπηρετηθεί το 68% των πολιτών. Από 3 έως 6 λεπτά είναι ο χρόνος που χρειάζεται ώστε να εξυπηρετηθεί το 68 + 13,5 = 81,5 % των πολιτών. Διάμεσος $\delta = 5'$, εύρος $R = 6s = 6'$. Είναι $CV = \dfrac{s}{\bar{x}} = \dfrac{1}{5} = \dfrac{2}{10} = 0,2$ ή $CV = 20\%$.

Εφαρμογή 42. Η ωριαία αποζημίωση (σε £) των υπαλλήλων δικηγορικής εταιρείας δίνεται από τον παρακάτω στατιστικό πίνακα. Αφού τον συμπληρώσετε, υπολογίστε τη μέση ωριαία αποζημίωση και τα μέτρα διασποράς.

Αμοιβή σε £ [...−...)	$F_i\%$	$f_i\%$	f_i	x_i	$x_i f_i$	x_i^2	$x_i^2 f_i$
20 – 30	$F_1\% = 20\ \%$	$f_1\% =$	$f_1 =$	$x_1 =$	$x_1 f_1 =$	$x_1^2 =$	$x_1^2 f_1 =$
30 – 40	$F_2\% = 45\ \%$	$f_2\% =$	$f_2 =$	$x_2 =$	$x_2 f_2 =$	$x_2^2 =$	$x_2^2 f_2 =$
40 – 50	$F_3\% = 85\ \%$	$f_3\% =$	$f_3 =$	$x_3 =$	$x_3 f_3 =$	$x_3^2 =$	$x_3^2 f_3 =$
50 – 60	$F_4\% = 95\ \%$	$f_4\% =$	$f_4 =$	$x_4 =$	$x_4 f_4 =$	$x_4^2 =$	$x_4^2 f_4 =$
60 – 70	$F_5\% = 100\ \%$	$f_5\% =$	$f_5 =$	$x_5 =$	$x_5 f_5 =$	$x_5^2 =$	$x_5^2 f_5 =$
Σύνολο		$\sum_{i=1}^{5} f_i\% =$	$\sum_{i=1}^{5} f_i =$		$\sum_{i=1}^{5} x_i f_i =$		$\sum_{i=1}^{5} x_i^2 f_i =$

Λύση.

Αμοιβή σε £ [...−...)	$F_i\%$	$f_i\%$	f_i	x_i	$x_i f_i$	x_i^2	$x_i^2 f_i$
20 – 30	$F_1\% = 20\ \%$	$f_1\% = 20\ \%$	$f_1 = 0{,}20$	$x_1 = 25$	$x_1 f_1 = 5$	$x_1^2 = 625$	$x_1^2 f_1 = 625 \cdot 0{,}2 = 62{,}5 \cdot 2$
30 – 40	$F_2\% = 45\ \%$	$f_2\% = 25\ \%$	$f_2 = 0{,}25$	$x_2 = 35$	$x_2 f_2 = 8{,}75$	$x_2^2 = 35^2$	$x_2^2 f_2 = 35^2 \cdot 0{,}25$
40 – 50	$F_3\% = 85\ \%$	$f_3\% = 40\ \%$	$f_3 = 0{,}40$	$x_3 = 45$	$x_3 f_3 = 18$	$x_3^2 = 45^2$	$x_3^2 f_3 = 45^2 \cdot 0{,}4$
50 – 60	$F_4\% = 95\ \%$	$f_4\% = 10\ \%$	$f_4 = 0{,}10$	$x_4 = 55$	$x_4 f_4 = 5{,}5$	$x_4^2 = 55^2$	$x_4^2 f_4 = 55^2 \cdot 0{,}1$
60 – 70	$F_5\% = 100\ \%$	$f_5\% = 5\ \%$	$f_5 = 0{,}05$	$x_5 = 65$	$x_5 f_5 = 3{,}25$	$x_5^2 = 65^2$	$x_5^2 f_5 = 65^2 \cdot 0{,}05$
Σύνολο		$\sum_{i=1}^{5} f_i\% = 100$	$\sum_{i=1}^{5} f_i = 1$		$\sum_{i=1}^{5} x_i f_i = 40{,}5$		$\sum_{i=1}^{5} x_i^2 f_i =$

$R = 70 - 20 = 50$ $\qquad \bar{x} = \sum_{i=1}^{5} x_i f_i = 40{,}5\ £$ $\qquad s^2 = \sum_{i=1}^{5} x_i^2 f_i - \left(\bar{x}\right)^2 = \sum_{i=1}^{5} x_i^2 f_i - 40{,}5^2 = \ldots$ $\qquad CV = \frac{s}{\bar{x}} = \frac{s}{40{,}5} = \ldots$

Εφαρμογή 43. Υπολογίστε τα μέτρα διασποράς για την ακόλουθη κατανομή βαθμών είκοσι μαθημάτων.

Κλάσεις $[\ldots-\ldots)$	ν_i	x_i	$x_i\nu_i$	x_i^2	f_i	$x_i^2 f_i$
$0-2$	$\nu_1=4$					$x_1^2 f_1=$
$2-4$	$\nu_2=5$					$x_2^2 f_2=$
$4-6$	$\nu_3=2$					$x_3^2 f_3=$
$6-8$	$\nu_4=5$					$x_4^2 f_4=$
$8-10$	$\nu_5=4$					$x_5^2 f_5=$
Σύνολο	$\nu=\sum_{i=1}^{5}\nu_i=$		$\sum_{i=1}^{5}x_i\nu_i=$		$\sum_{i=1}^{5}f_i=$	$\sum_{i=1}^{5}x_i^2 f_i=$

Λύση.

Εύρος $R=10-0=10$,

Μέση τιμή $\bar{x}=\dfrac{\sum_{i=1}^{5}x_i\nu_i}{\nu}=\dfrac{100}{20}=5$,

Διασπορά $s^2=\overline{x^2}-\left(\bar{x}\right)^2=\sum_{i=1}^{5}x_i^2 f_i-5^2=33,4-25=8,4$

Τυπική απόκλιση $s=\sqrt{s^2}=\sqrt{8,4}=2,9$

Συντελεστής μεταβολής $CV=\dfrac{s}{\bar{x}}=\dfrac{2,9}{5}=\dfrac{5,8}{10}=0,58$

Άρα, $CV\%=58\%$. Όχι ομοιογενές.

Κλάσεις $[\ldots-\ldots)$	ν_i	x_i	$x_i\nu_i$	x_i^2	f_i	$x_i^2 f_i$
$0-2$	$\nu_1=4$	$x_1=1$	$x_1\nu_1=4$	$x_1^2=1$	$f_1=0,2$	$x_1^2 f_1=0,2$
$2-4$	$\nu_2=5$	$x_2=3$	$x_2\nu_2=15$	$x_2^2=9$	$f_2=0,25$	$x_2^2 f_2=2,25$
$4-6$	$\nu_3=2$	$x_3=5$	$x_3\nu_3=10$	$x_3^2=25$	$f_3=0,1$	$x_3^2 f_3=2,5$
$6-8$	$\nu_4=5$	$x_4=7$	$x_4\nu_4=35$	$x_4^2=49$	$f_4=0,25$	$x_4^2 f_4=12,25$
$8-10$	$\nu_5=4$	$x_5=9$	$x_5\nu_5=36$	$x_5^2=81$	$f_5=0,2$	$x_5^2 f_5=16,2$
Σύνολο	$\nu=\sum_{i=1}^{5}\nu_i=20$		$\sum_{i=1}^{5}x_i\nu_i=100$		$\sum_{i=1}^{5}f_i=1$	$\sum_{i=1}^{5}x_i^2 f_i=33,4$

Εφαρμογή 44. Για την ακόλουθη κατανομή του βάρους 40 containers, βρείτε επικρατούσα τιμή, αριθμητικό μέσο, διάμεσο, εύρος, διασπορά και τυπική απόκλιση.

x_i	ν_i	$x_i\nu_i$
$x_1=1$	$\nu_1=6$	
$x_2=2$	$\nu_2=7$	
$x_3=3$	$\nu_3=8$	
$x_4=4$	$\nu_4=9$	
$x_5=5$	$\nu_5=10$	
Σύνολο	$\nu=\sum_{i=1}^{5}\nu_i=40$	$\sum_{i=1}^{5}x_i\nu_i=$

Λύση.

x_i	ν_i	$x_i\nu_i$
$x_1=1$	$\nu_1=6$	$x_1\nu_1=6$
$x_2=2$	$\nu_2=7$	$x_2\nu_2=14$
$x_3=3$	$\nu_3=8$	$x_3\nu_3=24$
$x_4=4$	$\nu_4=9$	$x_4\nu_4=36$
$x_5=5$	$\nu_5=10$	$x_5\nu_5=50$
Σύνολο	$\nu=\sum_{i=1}^{5}\nu_i=40$	$\sum_{i=1}^{5}x_i\nu_i=120$

Επικρατούσα τιμή είναι αυτή με τη μεγαλύτερη συχνότητα, δηλαδή η $x_5=5$.

Αριθμητικός μέσος (ή μέση τιμή) είναι ο $\bar{x}=\dfrac{\sum_{i=1}^{6}x_i\nu_i}{\nu}=\dfrac{120}{40}=3$.

Διάμεσος των 40 παρατηρήσεων είναι η παρατήρηση $x_3=3$.

Εύρος ή διακύμανση είναι $R=5-1=4$.

Διασπορά είναι $s^2=\dfrac{\sum_{i=1}^{5}\left(x_i-\bar{x}\right)^2\cdot\nu_i}{\nu}=$

$$\frac{\left(x_1-\bar{x}\right)^2\cdot\nu_1+\left(x_2-\bar{x}\right)^2\cdot\nu_2+\left(x_3-\bar{x}\right)^2\cdot\nu_3+\left(x_4-\bar{x}\right)^2\cdot\nu_4+\left(x_5-\bar{x}\right)^2\cdot\nu_5}{\nu}=$$

$$\frac{(1-3)^2\cdot 6+(2-3)^2\cdot 7+(3-3)^2\cdot 8+(4-3)^2\cdot 9+(5-3)^2\cdot 10}{40}=$$

$$\frac{2^2\cdot 6+1^2\cdot 7+0^2\cdot 8+1^2\cdot 9+2^2\cdot 10}{40}=\frac{24+7+9+40}{40}=\frac{80}{40}=2.$$

Τυπική απόκλιση είναι $s=\sqrt{s^2}=\sqrt{2}\cong 1,4$.

Συντελεστής μεταβολής είναι $CV=\dfrac{s}{\bar{x}}=\dfrac{\sqrt{2}}{3}\cong\dfrac{1,4}{3}\cong 0,466$.

Άρα $CV\%\cong 46,6\%$, συνεπώς όχι ομοιογενές.

Εφαρμογή 45. Οι εισπράξεις (σε χιλιάδες €) ενός δείγματος δέκα υποκαταστημάτων εμπορικής επιχειρήσεως, σε ένα μήνα φαίνονται στον παρακάτω στατιστικό πίνακα. Αφού τον συμπληρώσετε, υπολογίστε επικρατούσα τιμή, διάμεσο, μέση τιμή $\bar{x}$ των εισπράξεων, εύρος, διασπορά, τυπική απόκλιση και συντελεστή μεταβολής.

Εισπράξεις (σε χιλιάδες €) x_i	Συχνότητα ν_i	f_i	$x_i-\bar{x}$	$\left(x_i-\bar{x}\right)^2$	$\nu_i\left(x_i-\bar{x}\right)^2$
$x_1=15$	$\nu_1=4$	$f_1=$			
$x_2=20$	$\nu_2=2$	$f_2=$			
$x_3=30$	$\nu_3=1$	$f_3=$			
$x_4=50$	$\nu_4=3$	$f_4=$			
Σύνολο	$\nu=\sum_{i=1}^{4}\nu_i=10$	$\sum_{i=1}^{4}f_i=$			$\sum_{i=1}^{4}\nu_i\left(x_i-\bar{x}\right)^2=$

Λύση.

$$\bar{x}=\frac{\sum_{i=1}^{4}\nu_i\cdot x_i}{\nu}=\frac{4\cdot 15+2\cdot 20+30+150}{10}=\frac{280}{10}=28$$

Επικρατούσα τιμή είναι η $x_1=15$. Διάμεσος είναι η τιμή $x_2=20$

Εύρος $R=50-15=35$

$$s^2=\frac{\sum_{i=1}^{4}\nu_i\left(x_i-\bar{x}\right)^2}{\nu}=\frac{2260}{10}=226$$

$$s=\sqrt{s^2}=\sqrt{226}\simeq 15$$

$$CV=\frac{s}{\bar{x}}\simeq\frac{15}{10}=1,5$$

Εισπράξεις (σε χιλιάδες €) x_i	Συχνότητα ν_i	f_i	$x_i - \overline{x}$	$\left(x_i - \overline{x}\right)^2$	$\nu_i\left(x_i - \overline{x}\right)^2$
$x_1 = 15$	$\nu_1 = 4$	$f_1 = 0,4$	$15 - 28 = -13$	169	$169 \cdot 4 = 676$
$x_2 = 20$	$\nu_2 = 2$	$f_2 = 0,2$	$20 - 28 = -8$	64	$64 \cdot 2 = 128$
$x_3 = 30$	$\nu_3 = 1$	$f_3 = 0,1$	$30 - 28 = 2$	4	$4 \cdot 1 = 4$
$x_4 = 50$	$\nu_4 = 3$	$f_4 = 0,3$	$50 - 28 = 22$	484	$484 \cdot 3 = 1452$
Σύνολο	$\nu = \sum_{i=1}^{4} \nu_i = 10$	$\sum_{i=1}^{4} f_i = 1$			$\sum_{i=1}^{4} \nu_i\left(x_i - \overline{x}\right)^2 = 2260$

Εφαρμογή 46. Υπολογίστε τα μέτρα διασποράς για τον παρακάτω στατιστικό πίνακα:

Κλάσεις	x_i	$F_i\%$	$f_i\%$	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f$
$[0,\ 2)$		$F_1 = 10$					
$[2,\ 4)$		$F_2 = 60$					
$[4,\ 6)$		$F_3 = 100$					
Σύνολο			$\sum_{i=1}^{3} f_i\% =$	$\sum_{i=1}^{3} f_i =$	$\sum_{i=1}^{3} x_i \cdot f_i =$		$\sum_{i=1}^{3} x_i^2 \cdot f_i =$

Λύση. Εύρος $R = 6 - 0 = 6$

Κλάσεις	x_i	$F_i\%$	$f_i\%$	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f$
$[0,\ 2)$	$x_1 = 1$	$F_1 = 10$	$f_1 = 10$	$f_1 = 0{,}1$	$x_1 f_1 = 0{,}1$	$x_1^2 = 1$	$x_1^2 f_1 = 0{,}1$
$[2,\ 4)$	$x_2 = 3$	$F_2 = 60$	$f_2 = 50$	$f_2 = 0{,}5$	$x_2 f_2 = 1{,}5$	$x_2^2 = 9$	$x_2^2 f_2 = 4{,}5$
$[4,\ 6)$	$x_3 = 5$	$F_3 = 100$	$f_3 = 40$	$f_3 = 0{,}4$	$x_3 f_3 = 2$	$x_3^2 = 25$	$x_3^2 f_3 = 10$
Σύνολο			$\sum_{i=1}^{3} f_i\% = 100$	$\sum_{i=1}^{3} f_i = 1$	$\sum_{i=1}^{3} x_i \cdot f_i = 3{,}6$		$\sum_{i=1}^{3} x_i^2 \cdot f_i = 14{,}6$

Υπολογισμός μέσης τιμής $\bar{x}$. $\bar{x} = \sum_{i=1}^{3} x_i \cdot f_i = 3{,}6$

Υπολογισμός συντελεστή μεταβολής. $CV = \frac{S}{\bar{x}} = \frac{1{,}28}{3{,}6} = \frac{12{,}8}{36} = \frac{6{,}4}{18} = \frac{3{,}2}{9} = 0{,}355$ Άρα $CV = 35{,}5\%$

Συνεπώς το δείγμα δεν είναι ομοιογενές.

1ος τρόπος υπολογισμού διασποράς, τυπικής αποκλίσεως.

$$s^2=\sum_{i=1}^{3}x_i^2\cdot f_i-\left(\overline{x}\right)^2=14,6-3,6^2=14,6-12,96=1,64$$ Άρα $s=\sqrt{s^2}=\sqrt{1,64}=1,28$

2[ος] τρόπος υπολογισμού διασποράς, τυπικής αποκλίσεως.

$s^2=\overline{x^2}-\left(\overline{x}\right)^2=14,6-12,96=1,64$ Άρα $s=\sqrt{s^2}=\sqrt{1,64}=1,28$

3[ος] τρόπος υπολογισμού διασποράς, τυπικής αποκλίσεως.

$$s^2=\frac{\sum_{i=1}^{3}\left(x_i-\overline{x}\right)^2\cdot\nu_i}{\nu}=\frac{\left(x_1-\overline{x}\right)^2\cdot\nu_1+\left(x_2-\overline{x}\right)^2\cdot\nu_2+\left(x_3-\overline{x}\right)^2\cdot\nu_3}{\nu}=$$

$$\frac{\left(x_1-\overline{x}\right)^2\cdot\nu_1}{\nu}+\frac{\left(x_2-\overline{x}\right)^2\cdot\nu_2}{\nu}+\frac{\left(x_3-\overline{x}\right)^2\cdot\nu_3}{\nu}=\left(x_1-\overline{x}\right)^2\cdot f_1+\left(x_2-\overline{x}\right)^2\cdot f_2+\left(x_3-\overline{x}\right)^2\cdot f_3=$$

$$\left(1-3,6\right)^2\cdot\frac{1}{10}+\left(3-3,6\right)^2\cdot\frac{5}{10}+\left(5-3,6\right)^2\cdot\frac{4}{10}=$$

$$\frac{2,6^2}{10}+\frac{0,6^2}{2}+\frac{1,4^2\cdot 2}{5}=0,676+0,18+0,784=1,64$$ Άρα, $s=\sqrt{s^2}=\sqrt{1,64}=1,28$

Εφαρμογή 47. Υπολογίστε τα μέτρα διασποράς.

x_i	ν_i
$x_1 = 1$	$\nu_1 = 5$
$x_2 = 2$	$\nu_2 = 35$
$x_3 = 3$	$\nu_3 = 10$
Σύνολο	$\nu = \sum_{i=1}^{3} \nu_i =$

Λύση. Εύρος $R = 3 - 1 = 2$ $\qquad \bar{x} = \dfrac{\sum_{i=1}^{3} \nu_i \cdot x_i}{\nu} = \dfrac{105}{50} = \dfrac{210}{100} = 2,1$

x_i	ν_i	$\nu_i \cdot x_i$
$x_1 = 1$	$\nu_1 = 5$	$\nu_1 \cdot x_1 = 5$
$x_2 = 2$	$\nu_2 = 35$	$\nu_2 \cdot x_2 = 70$
$x_3 = 3$	$\nu_3 = 10$	$\nu_3 \cdot x_3 = 30$
Σύνολο	$\nu = \sum_{i=1}^{3} \nu_i = 50$	$\sum_{i=1}^{3} \nu_i \cdot x_i = 105$

$$s^2 = \frac{\sum_{i=1}^{3} \nu_i \left(x_i - \bar{x}\right)^2}{\nu} = \frac{5\left(x_1 - \bar{x}\right)^2 + 35\left(x_2 - \bar{x}\right)^2 + 10\left(x_3 - \bar{x}\right)^2}{50} =$$

$$\frac{5(1-2,1)^2 + 35(2-2,1)^2 + 10(3-2,1)^2}{50} = \frac{14,9}{50} = \frac{29}{100} = 0,29 .$$

Τυπική απόκλιση $s = \sqrt{s^2} = \sqrt{0,29} = \dfrac{\sqrt{29}}{\sqrt{100}} = \dfrac{\sqrt{29}}{10} \cong \dfrac{5,3}{10} = 0,53$.

Συντελεστής μεταβολής $CV = \dfrac{s}{\bar{x}} \simeq \dfrac{0,53}{0,29} = \dfrac{53}{29}$.

Εφαρμογή 48. Βρείτε τις τιμές των συχνοτήτων ν_1, ν_2, ν_3 της μεταβλητής X αν $\overline{x}=1,5$ και $s^2=4,5$, για τα δεδομένα του παρακάτω πίνακα.

x_i	ν_i
$x_1=0$	ν_1
$x_2=2$	ν_2
$x_3=5$	ν_3
Σύνολο	80

Λύση. • Είναι $\nu_1+\nu_2+\nu_3=80$.

• Είναι $\overline{x}=\dfrac{\sum_{i=1}^{3}x_i\nu_i}{\nu}=\dfrac{x_1\nu_1+x_2\nu_2+x_3\nu_3}{80}=\dfrac{0\nu_1+2\nu_2+5\nu_3}{80}=\dfrac{2\nu_2+5\nu_3}{80}$

Από $\overline{x}=1,5$ έπεται ότι $\dfrac{2\nu_2+5\nu_3}{80}=1,5\Leftrightarrow 2\nu_2+5\nu_3=120$.

• Από $s^2=\overline{x^2}-\left(\overline{x}\right)^2\Leftrightarrow 4,5=\overline{x^2}-(1,5)^2\Leftrightarrow 4,5=\overline{x^2}-2,25\Leftrightarrow \overline{x^2}=6,75$

Και $\overline{x^2}=\sum_{i=1}^{3}x_i^2f_i=x_1^2f_1+x_2^2f_2+x_3^2f_3=0f_1+4f_2+25f_3=4f_2+25f_3=$

$4\dfrac{\nu_2}{\nu}+25\dfrac{\nu_3}{\nu}=4\dfrac{\nu_2}{80}+25\dfrac{\nu_3}{80}$ προκύπτει ότι $4\dfrac{\nu_2}{80}+25\dfrac{\nu_3}{80}=6,75\Leftrightarrow 4\nu_2+25\nu_3=540$.

Συνεπώς $\left\{\begin{array}{l}\nu_1+\nu_2+\nu_3=80\\ 2\nu_2+5\nu_3=120\\ 4\nu_2+25\nu_3=540\end{array}\right\}\Leftrightarrow\left\{\begin{array}{l}\nu_1=50\\ \nu_2=10\\ \nu_3=20\end{array}\right.$.

Stem and leaf plots (φυλλογράφημα, μίσχος και φυλλαράκια).

Έστω ότι έχω τις τιμές

136,4	110,9	120	110,1	110,6
116,2	99	134,5	121,7	

Ζητείται να παρασταθούν.

Μέθοδος εργασίας.

1. Οι αριθμοί στρογγυλοποιούνται στον πλησιέστερο ακέραιο.

136	111	120	110	111
116	99	135	122	

2. Επιλέγω πρώτα stem (μίσχοι ή οδηγούνταν ψηφία) ΔΕΚΑΔΕΣ και μετά leaves (φυλλαράκια ή επόμενα ψηφία) ΜΟΝΑΔΕΣ.

stems	leaves
13	6
11	1
12	0
11	0
11	1
11	6
9	9
13	5
12	2

3. Καταγράφομε τα stems.

4. Διατάσσομε τα stems κατά αύξουσα τάξη, γράφοντας τα κατακόρυφα.

Stem and leaf plot.

Δεκάδες	Μονάδες
9	9
10	--
11	1, 0, 1, 6
12	0, 2
13	6, 5

5. Γράφομε τα leaves στην ίδια γραμμή που βρίσκεται το αντίστοιχο τους stem.

6. Ελέγχω αν έχω καταγράψει όλα τα φύλλα. Ο αριθμός τους πρέπει να ισούται με το σύνολο των παρατηρήσεων.

Παρατήρηση. Αν το δω ανάποδα, είναι διάγραμμα συχνοτήτων. Το φυλλογράφημα παρουσιάζει την ίδια οπτική εικόνα με το διάγραμμα των συχνοτήτων, αλλά οι αριθμοί εδώ διατηρούνται.

Το IQ test.

Οι έννοιες του «δείκτη νοημοσύνης» και των test ευφυΐας, εισέρχονται όλο και περισσότερο στην καθημερινή μας ζωή, με σκοπό την επιλογή του κατάλληλου ανθρώπου για την κατάλληλη θέση. Ο όρος «δείκτης νοημοσύνης» αποτελεί την απόδοση στα ελληνικά του αγγλικού όρου « Intelligence Quotient » (IQ) που ερμηνεύεται ως πηλίκο ευφυΐας ή διανοητικό πηλίκο. Ο όρος αυτός υιοθετήθηκε στην επιστημονική ορολογία το 1912 και από τότε γνώρισε μεγάλη δημοσιότητα. Εκφράζει το πηλίκο της «διανοητικής ηλικίας» (Mental Age – M.A.) προς τη «χρονολογική ηλικία» (Chronological Age– C.A.) πολλαπλασιασμένο επί 100.

$$\text{I.Q.}=\frac{\text{M.A.}}{\text{C.A.}}100$$

Το M.A. εξαρτάται από την απόδοση του ατόμου σε μία αναγνωρισμένη δοκιμασία ευφυΐας (I.Q. test). Το C.A. εκφράζει την πραγματική ηλικία του ατόμου.

Αν άτομο ηλικίας 10 ετών επιτύχει σε test ευφυΐας 120 βαθμούς, έχει M.A. = 12, δηλαδή νοητική ηλικία δωδεκάχρονου παιδιού. Η ανάπτυξη της ευφυΐας ολοκληρώνεται περί τα 15 έως 17 έτη της ηλικίας του ατόμου. Κατά άλλους ειδικούς η νοητική ανάπτυξη ολοκληρώνεται στα 16 έτη, οπότε τίθεται ως συμβατικό επίπεδο η ηλικία των 15,5 ετών. Συνεπώς δεν έχει νόημα να θεωρηθεί ότι ένα άτομο ηλικίας 22 ετών έχει διανοητική ηλικία 28, καθόσον η διανοητική ηλικία είναι ίδια και στις δυο περιπτώσεις.

Σε ενήλικα άτομα του πληθυσμού, το IQ test χρησιμοποιείται για τον προσδιορισμό της θέσεως τους στην κλίμακα κατανομής του IQ και όχι για τον προσδιορισμό της διανοητικής ηλικίας του εξεταζόμενου ατόμου. Προϊόντος του χρόνου, ο παρονομαστής του ανωτέρω τύπου αυξάνεται, ενώ ο αριθμητής παραμένει σταθερός, συνεπώς παρατηρείται διαρκής και σταθερή μείωση του δείκτη νοημοσύνης. Προς τούτο για όλα τα άτομα ηλικίας άνω των 16 ετών, θεωρούμε ότι C.A. = 16 ή C.A. = 15,5.

Τα σύγχρονα IQ test εξετάζουν ένα δείγμα ικανοτήτων που θεωρείται ότι σχετίζονται με την ευφυΐα. Το test Stantord – Binet και οι παραλλαγές του, εξετάζουν συστηματικά στοιχεία ευφυΐας ατόμων δυτικών κοινωνιών. Τα test λεκτικών ικανοτήτων, οι κλίμακες του Raven και τα test πρακτικής σκέψεως εξετάζουν διαφορετικά στοιχεία ευφυΐας το κάθε ένα. Οι βαθμολογίες που επιτυγχάνονται στα test αυτά είναι παραπλήσιες, αλλά όχι ταυτόσημες. Οι «Νοομομετρικής Προϊούσης Δυσχέρειας Κλίμακες του Raven» πλεονεκτούν στο ότι μετρούν ορισμένους θεμελιώδεις συντελεστές νοημοσύνης, ανεξάρτητα από το μορφωτικό επίπεδο του εξεταζομένου ατόμου.

Μπορεί να δοθεί σε άτομα διαφορετικών κοινωνικών, πολιτιστικών συστημάτων χωρίς να υπάρχουν σημαντικά σφάλματα κατά τη σύγκριση των αποτελεσμάτων, δηλαδή σφάλματα οφειλόμενα στη διαφορετική επίδραση που ασκούν οι διάφορες μορφές του πολιτισμού στην εξέλιξη της ευφυΐας του ατόμου.

Μπορεί να δοθεί σε αγράμματα άτομα, ή με δυσκολίες στην ακοή ή στην ομιλία, μιας και οι προφορικές υποδείξεις περιορίζονται στο ελάχιστο.

Οι κλίμακες του Raven δεν εξετάζουν σφαιρικά την ευφυΐα, αλλά επιλεκτικά ορισμένα μόνο στοιχεία νοημοσύνης, όπως αναλυτική ικανότητα, αριθμητική ικανότητα, χωροαντιληπτική ικανότητα, αίσθηση συμμετρίας, ικανότητα συσχετισμού σχημάτων και συμβόλων. Τα περισσότερα test νοημοσύνης έχουν χρονικό περιορισμό, συνήθως 30 έως 45 λεπτών και η κάθε δοκιμασία είναι δυσκολότερη από την προηγούμενη.

Ενδεικτικά, για test διάρκειας 30 λεπτών, όσοι το ολοκληρώσουν σε λιγότερο από 15 λεπτά πριμοδοτούνται με 5 επιπλέον βαθμούς, όσοι τελειώσουν σε λιγότερο από 20 λεπτά πριμοδοτούνται με 3 επιπλέον βαθμούς, ενώ όσοι χρειασθούν λιγότερο από 25 λεπτά πριμοδοτούνται με 2 επιπλέον βαθμούς.

Δεν αποτελούν τέλειες διαδικασίες και ως εκ τούτου δεν πρέπει να υπερτιμώνται. Αν η ευφυΐα δεν συνοδεύεται από άλλες ιδιότητες, πιθανό να αποδειχθεί επιζήμια και καταστροφική για το άτομο και το κοινωνικό σύνολο.

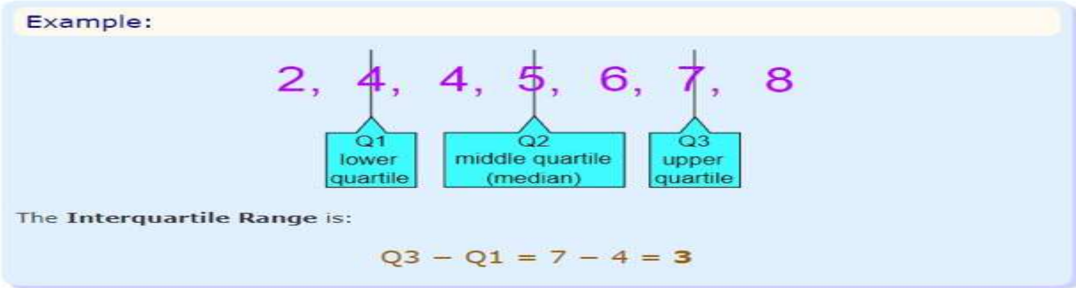
Τα ειδικά test για παιδιά 5 έως 11 ετών, έχουν περιορισμένη προγνωστική αξία. Όσο μικρότερη είναι η ηλικία του παιδιού, τόσο λιγότερο αξιόπιστα είναι τα αποτελέσματα των IQ test. Αν η βαθμολογία είναι κατά πολύ χαμηλότερη από το μέσο όρο, ενδείκνυται η παροχή βοήθειας από εξειδικευμένο παιδοψυχολόγο. Η διάγνωση της πνευματικής καθυστερήσεως δε μπορεί να στηριχθεί μόνο στα αποτελέσματα ενός IQ test, ειδικά αν πρόκειται για παιδιά.

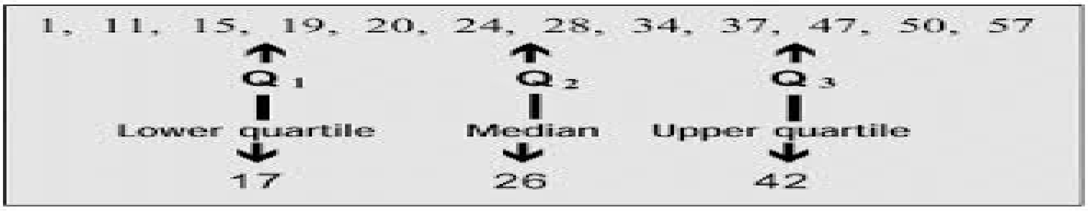
Η MENSA είναι η διεθνής λέσχη των ατόμων με εξαιρετικά υψηλό δείκτη νοημοσύνης, ιδρύθηκε το 1945 και αριθμεί πάνω από 45.000 μέλη παγκοσμίως.

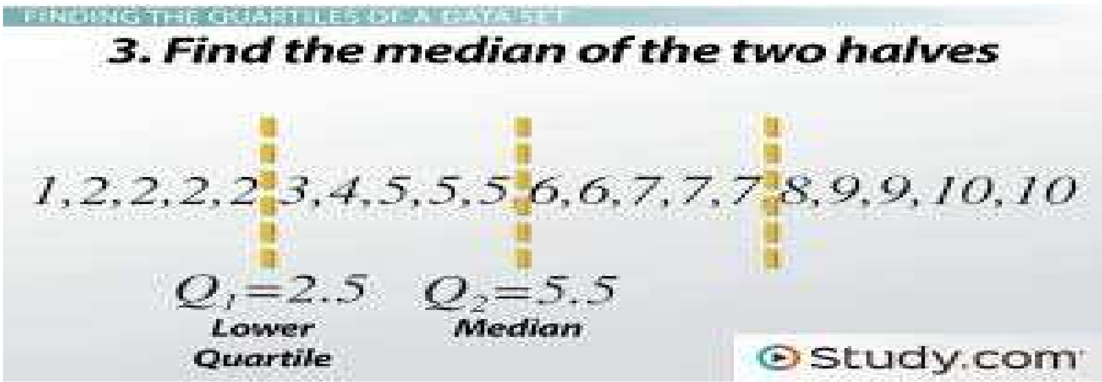
Δέχεται άτομα ανεξαρτήτως φυλής, θρησκείας, χρώματος, κοινωνικής, οικονομικής καταστάσεως, επαγγέλματος, μορφωτικού επιπέδου, αρκεί το IQ τους να ανήκει στο ανώτερο 2% του πληθυσμού (132 ή 133 βαθμοί σε ένα κλασσικό test).

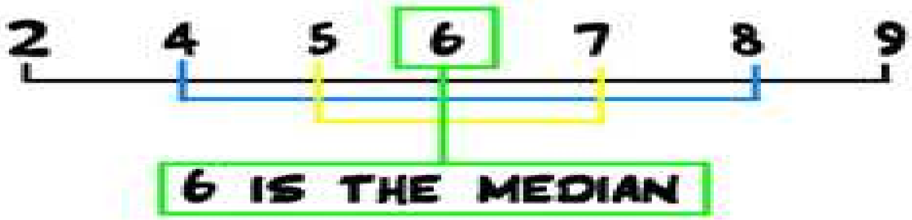
IQ	**Επίπεδο ευφυΐας**
Κάτω από 70	Πνευματικά υπολειπόμενος
70 – 79	Οριακό επίπεδο
80 – 89	Κάτω του μετρίου
90 – 109	Μέσης ευφυΐας
110 – 119	Άνω του μετρίου
120 – 140	Εξαιρετικής ευφυΐας
Πάνω από 140	Εξαιρετικά υψηλής ευφυΐας

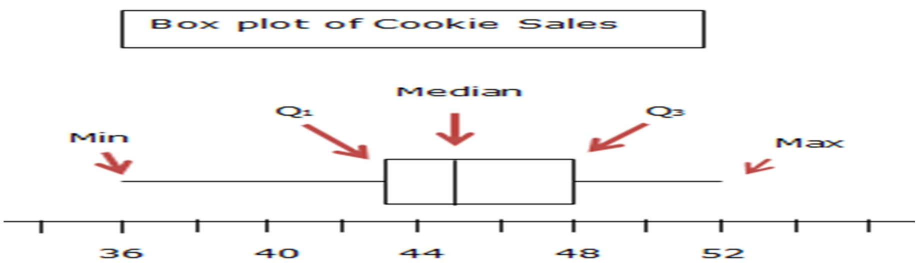
Θηκόγραμμα (boxplot) & Τεταρτημόρια (Quartiles)

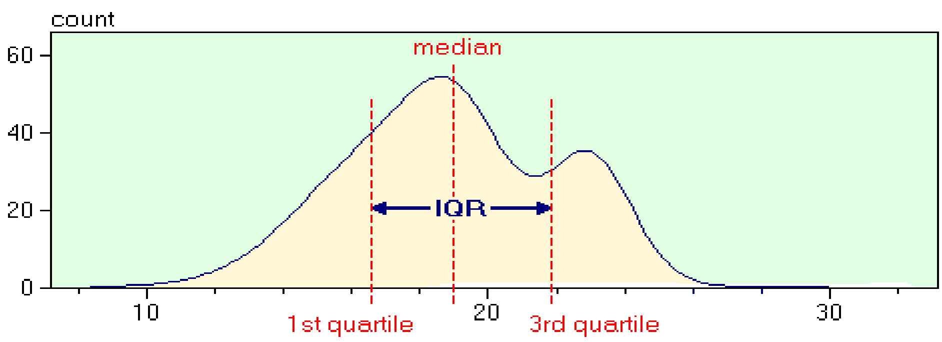


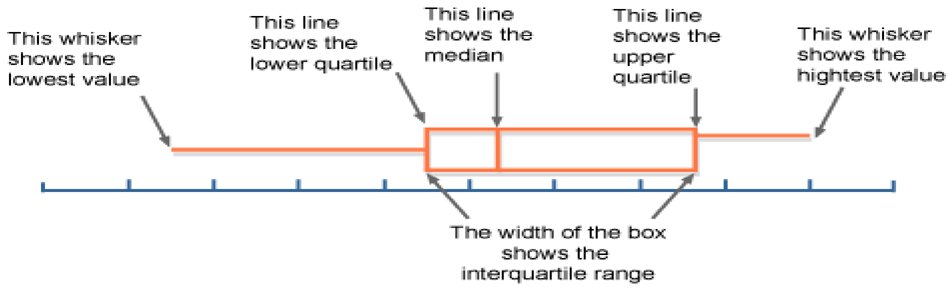


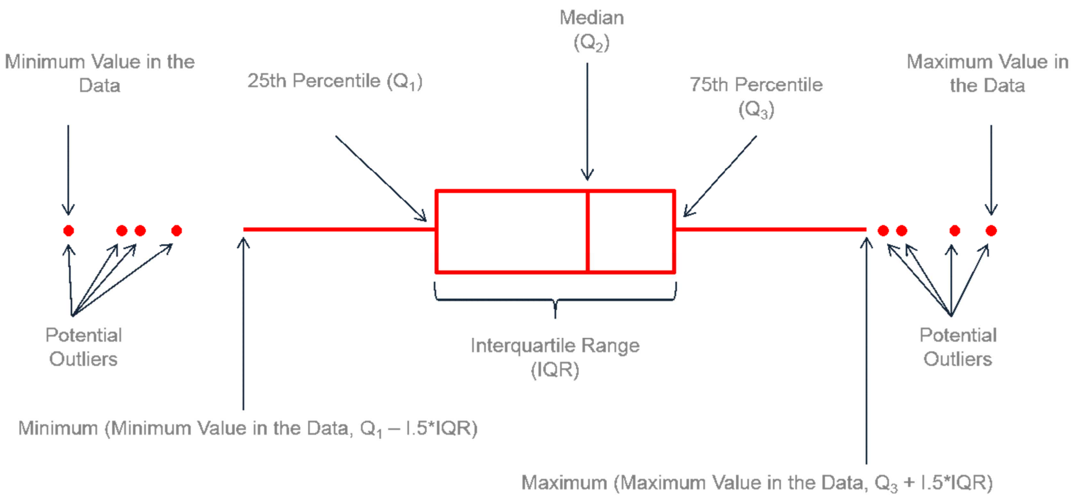


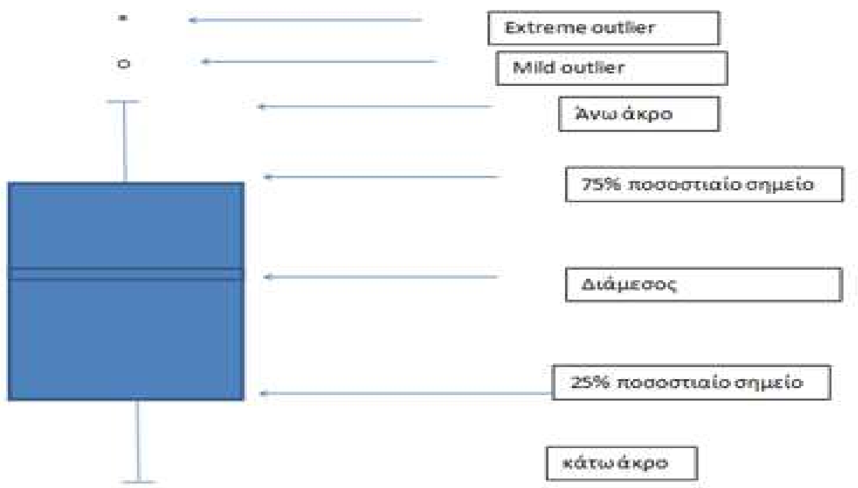


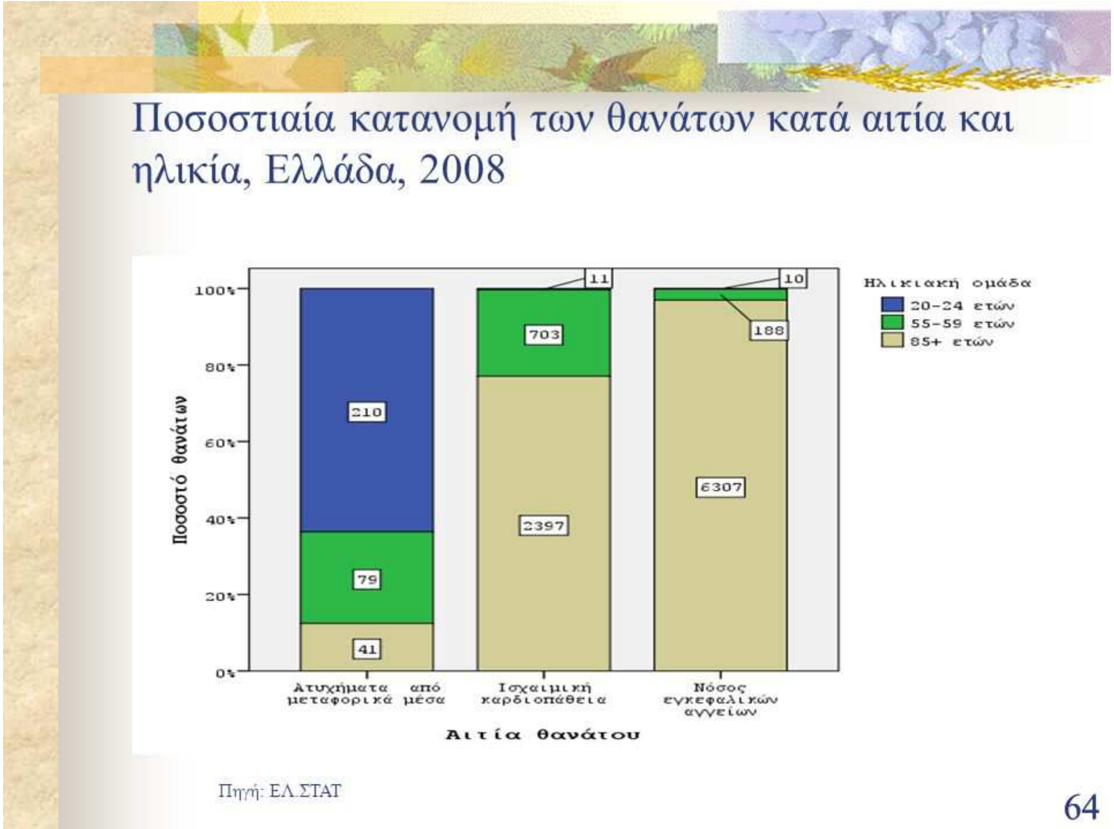


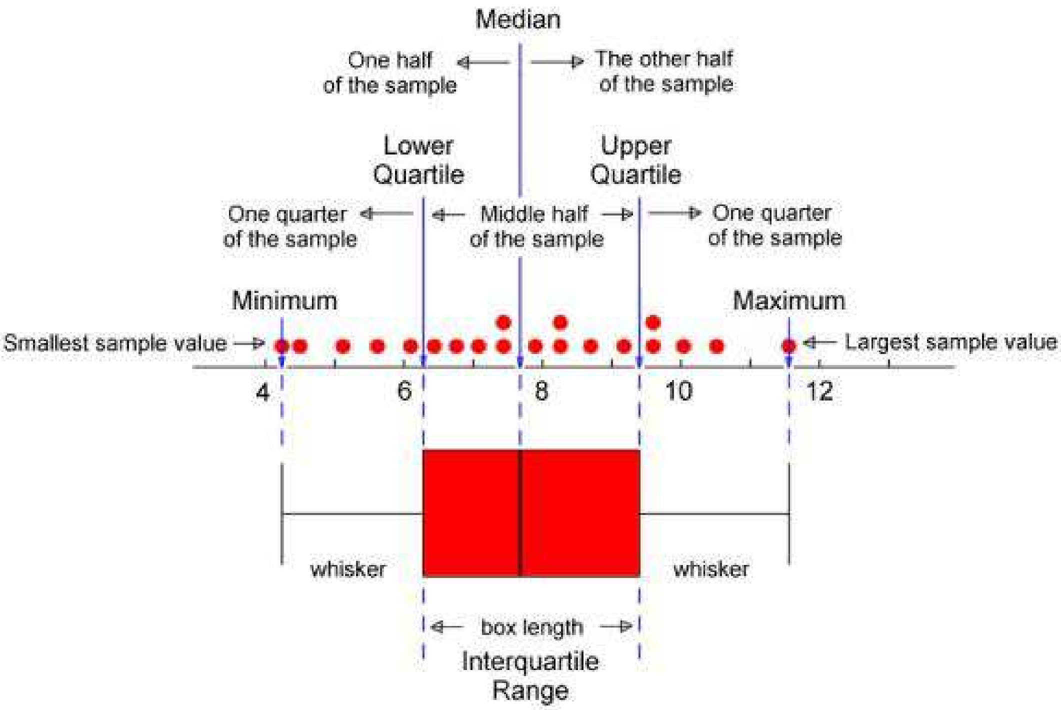

count
60
40
20
0
median
IQR
10
1st quartile
20
3rd quartile
30

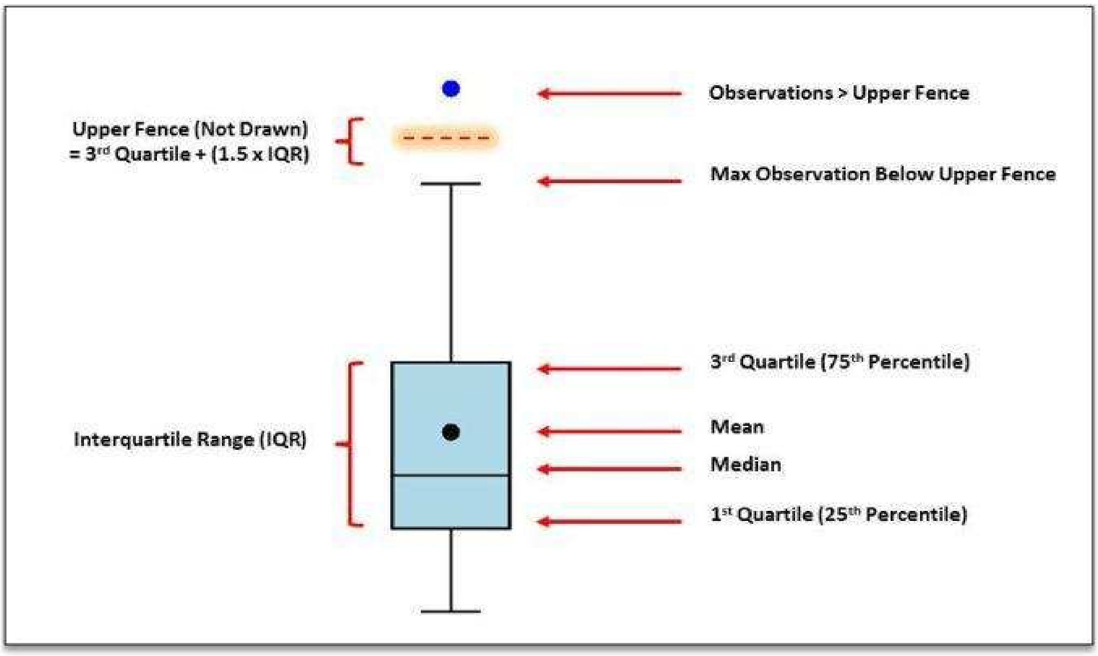

This whisker shows the lowest value
This line shows the lower quartile
This line shows the median
This line shows the upper quartile
This whisker shows the hightest value
The width of the box shows the interquartile range


Median (Q2)
Minimum Value in the Data
25th Percentile (Q1)
75th Percentile (Q3)
Maximum Value in the Data
Potential Outliers
Interquartile Range (IQR)
Potential Outliers
Minimum (Minimum Value in the Data, Q1 – 1.5*IQR)
Maximum (Maximum Value in the Data, Q3 + 1.5*IQR)

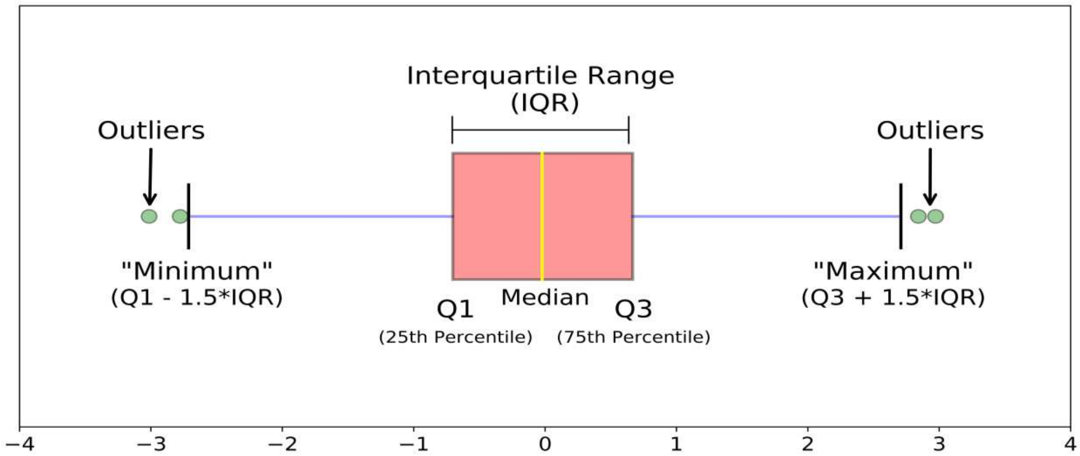

Extreme outlier
Mild outlier
Άνω άκρο
75% ποσοστιαίο σημείο
Διάμεσος
25% ποσοστιαίο σημείο
κάτω άκρο


Ποσοστιαία κατανομή των θανάτων κατά αιτία και ηλικία, Ελλάδα, 2008
Ποσοστό θανάτων
100%
80%
60%
40%
20%
0%
210
79
41
11
703
2397
10
188
6307
Ατυχήματα από μεταφορικά μέσα
Ισχαιμική καρδιοπάθεια
Νόσος εγκεφαλικών αγγείων
Αιτία θανάτου
Ηλικιακή ομάδα
20-24 ετών
55-59 ετών
85+ ετών
Πηγή: ΕΛ.ΣΤΑΤ
64

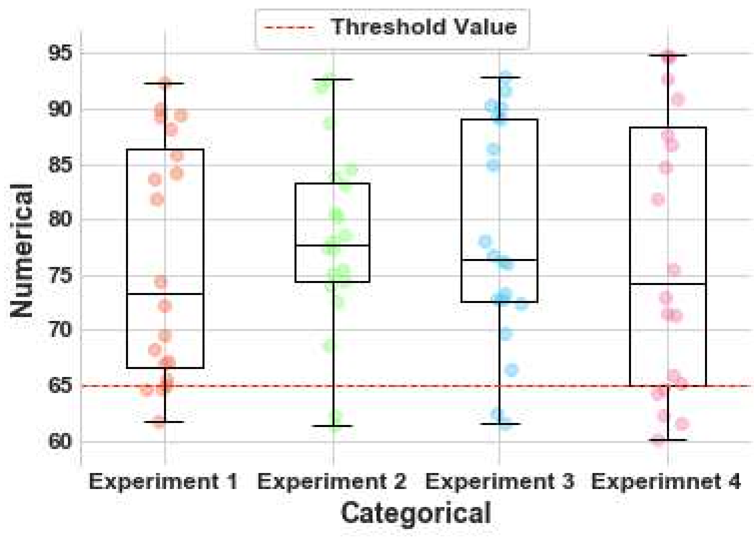

Median
One half of the sample
The other half of the sample
Lower Quartile
Upper Quartile
One quarter of the sample
Middle half of the sample
One quarter of the sample
Minimum
Maximum
Smallest sample value
Largest sample value
4
6
8
10
12
whisker
whisker
box length
Interquartile Range



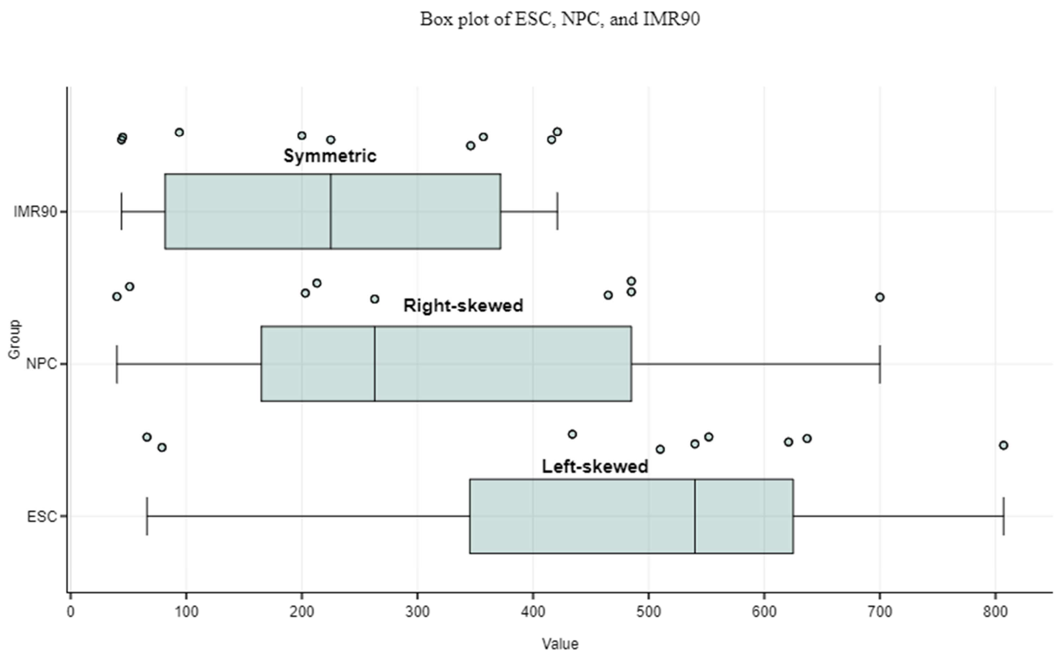
Observations > Upper Fence
Upper Fence (Not Drawn)
= 3rd Quartile + (1.5 x IQR)
Max Observation Below Upper Fence
3rd Quartile (75th Percentile)
Interquartile Range (IQR)
Mean
Median
1st Quartile (25th Percentile)

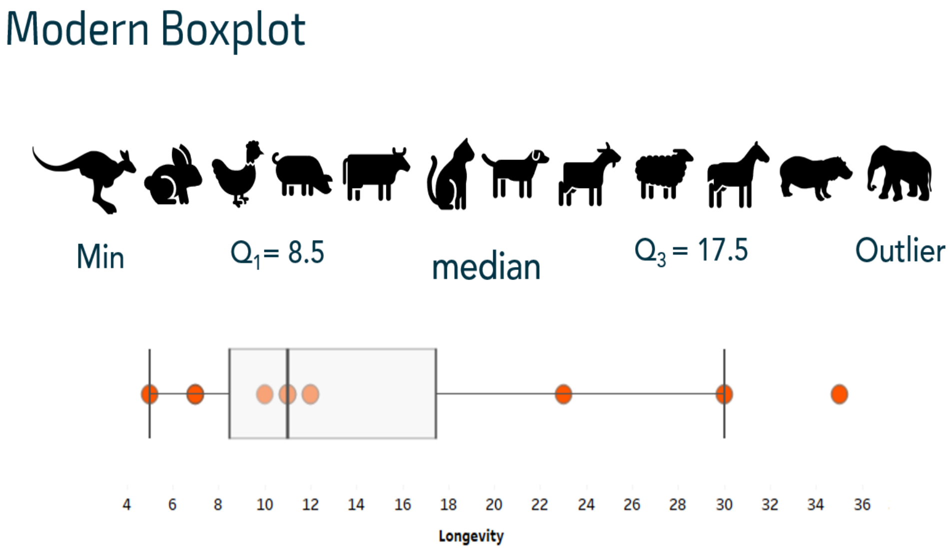


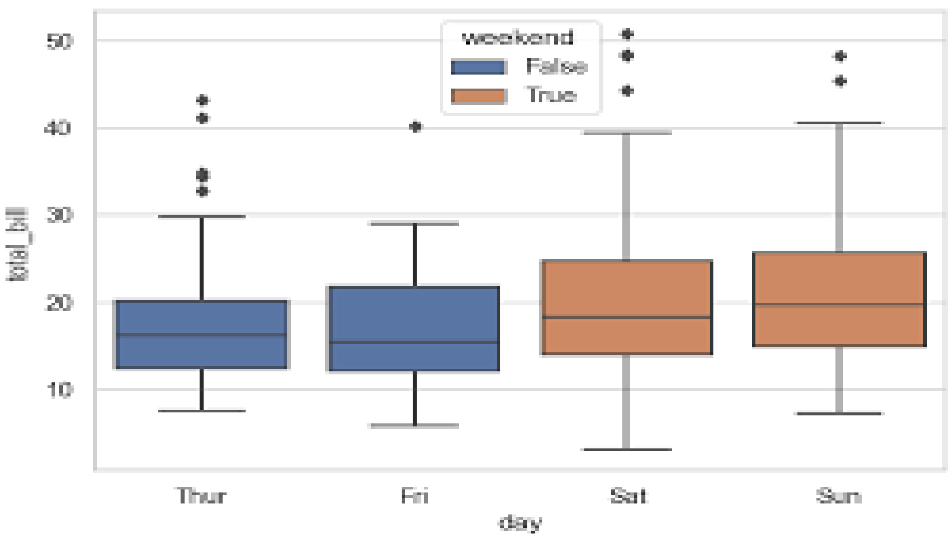
Interquartile Range
(IQR)
Outliers
Outliers
"Minimum"
(Q1 - 1.5*IQR)
Q1
Median
Q3
(25th Percentile)
(75th Percentile)
"Maximum"
(Q3 + 1.5*IQR)
−4
−3
−2
−1
0
1
2
3
4



Threshold Value
95
90
85
80
75
70
65
60
Numerical
Experiment 1
Experiment 2
Experiment 3
Experimnet 4
Categorical


Box plot of ESC, NPC, and IMR90
Symmetric
Right-skewed
Left-skewed
IMR90
NPC
ESC
Group
0
100
200
300
400
500
600
700
800
Value


Modern Boxplot
Min
Q_1= 8.5
median
Q_3 = 17.5
Outlier
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
Longevity


weekend
False
True
50
40
30
20
10
total_bill
Thur
Fri
Sat
Sun
day

Ερωτήσεις σωστού– λάθους.

1. Οι ποιοτικές μεταβλητές διακρίνονται σε συνεχείς και διακριτές.

2. Οι τιμές μίας ποιοτικής μεταβλητής είναι αριθμοί.

3. Το άθροισμα όλων των συχνοτήτων ν_i, των τιμών x_i μίας μεταβλητής X, ισούται με 1.

4. Το άθροισμα όλων των σχετικών συχνοτήτων f_i, των τιμών x_i μίας μεταβλητής X, ισούται με το μέγεθος ν του δείγματος.

5. Το άθροισμα όλων των σχετικών συχνοτήτων f_i, των τιμών x_i μίας μεταβλητής X, ισούται με 100.

6. Για τη σχετική συχνότητα f_i ισχύει ότι $f_i > 1 \ \ \forall i = 1, \ 2, ..., \ k$.

7. Οι αθροιστικές συχνότητες N_i εκφράζουν το πλήθος των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής x_i.

8. Οι αθροιστικές σχετικές συχνότητες F_i εκφράζουν το ποσοστό των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής x_i.

9. Η συχνότητα ν_i της τιμής x_i, μίας μεταβλητής X, είναι αρνητικός αριθμός.

10. Στην περίπτωση των <u>ποσοτικών</u> μεταβλητών, εκτός από τις συχνότητες ν_i & f_i, χρησιμοποιούνται και οι αθροιστικές συχνότητες N_i & F_i.

11. Το ραβδόγραμμα χρησιμοποιείται για τη γραφική παράσταση των τιμών μίας <u>ποιοτικής</u> μεταβλητής.

12. Το κυκλικό διάγραμμα (ή πίττα) χρησιμοποιείται για τη γραφική παράσταση των τιμών μίας ποιοτικής ή και ποσοτικής μεταβλητής.

13. Το ραβδόγραμμα αποτελείται από ορθογώνιες στήλες που οι βάσεις τους έχουν ίσα μήκη.

14. Η μέση τιμή, ενός συνόλου ν παρατηρήσεων, είναι μέτρο θέσεως.

15. Η διάμεσος δ, ενός δείγματος, είναι μέτρο διασποράς.

16. Αν οι παρατηρήσεις εκφράζονται σε cm, τότε και η διακύμανση εκφράζεται σε cm.

17. Ο συντελεστής μεταβολής (ή μεταβλητότητας) CV είναι ανεξάρτητος από τις μονάδες μετρήσεως.

18. Τα μέτρα διασποράς εκφράζουν τις αποκλίσεις των τιμών μίας μεταβλητής γύρω από τα μέτρα κεντρικής τάσεως.

19. Ο CV παραστάνει ένα μέτρο απόλυτης διασποράς και όχι σχετικής διασποράς.

20. Το εύρος R, ενός δείγματος ν παρατηρήσεων, ορίζεται ως το άθροισμα της μεγαλύτερης και της ελάχιστης παρατηρήσεως.

21. Το εύρος R, ενός δείγματος, εξαρτάται από τις δυο ακραίες παρατηρήσεις και είναι αξιόπιστο μέτρο διασποράς.

22. Τα μέτρα ασυμμετρίας καθορίζουν τη μορφή της κατανομής.

23. Τα μέτρα ασυμμετρίας εκφράζονται μόνο ως συνάρτηση των μέτρων θέσεως.

24. Τα μέτρα διασποράς εκφράζουν πόσο επεκτείνονται οι παρατηρήσεις γύρω από το «κέντρο» τους.

25. Η διασπορά είναι ο μέσος όρος των τετραγώνων των αποκλίσεων των t_i παρατηρήσεων από τη μέση τιμή τους $\overline{x}$.

26. Ο CV εκφράζει τη μεταβλητότητα των δεδομένων απαλλαγμένη από την επίδραση της μέσης τιμής.

27. Το κέντρο, κάθε κλάσεως ενός δείγματος, ισούται με την ημιδιαφορά των άκρων της κλάσεως.

28. Το εύρος του δείγματος χρησιμοποιείται για την κατασκευή ισοπλατών κλάσεων.

29. Η κατανομή συχνοτήτων με «κωδωνοειδή» μορφή ονομάζεται κανονική κατανομή.

30. Σε όλες τις περιπτώσεις, οι κλάσεις ενός δείγματος έχουν όλες το ίδιο πλάτος.

31. Όταν το πλήθος των τιμών μίας μεταβλητής είναι αρκετά μεγάλο, είναι απαραίτητο να ταξινομηθούν τα δεδομένα σε κλάσεις.

Ερωτήσεις.

1. Ποιες μεταβλητές ονομάζονται ποσοτικές;

2. Πότε, μία ποσοτική μεταβλητή, ονομάζεται διακριτή και πότε συνεχής;

3. Πότε ένα δείγμα ονομάζεται αντιπροσωπευτικό;

4. Τι ονομάζεται πληθυσμός και τι δείγμα;

5. Τι ονομάζεται μεταβλητή X ενός πληθυσμού και πως συμβολίζεται;

6. Ορισμός συχνότητας (ή απόλυτης συχνότητας) ν_i μίας τιμής x_i, της μεταβλητής X, ενός δείγματος μεγέθους ν.

7. Τι ονομάζεται αθροιστική συχνότητα N_i μίας τιμής x_i, της μεταβλητής X, ενός δείγματος μεγέθους ν;

8. Τι ονομάζεται σχετική συχνότητα f_i μίας τιμής x_i, της μεταβλητής X, ενός δείγματος μεγέθους ν και τι τιμές λαμβάνει;

9. Τι ονομάζεται επί τοις εκατό σχετική συχνότητα $f_i\%$ μίας τιμής x_i, της μεταβλητής X, ενός δείγματος μεγέθους ν και τι τιμές λαμβάνει;

10. Τι ονομάζεται επί τοις εκατό αθροιστική σχετική συχνότητα $F_i\%$ μίας τιμής x_i, της μεταβλητής X, ενός δείγματος μεγέθους ν και τι τιμές λαμβάνει;

11. Τι ονομάζεται πίνακας κατανομής συχνοτήτων;

12. Για τη γραφική παράσταση ποιού φαινομένου χρησιμοποιείται το χρονόγραμμα;

13. Σε μία κανονική κατανομή είναι $\bar{x}=30$ και $s=3$. Ποιο είναι, κατά προσέγγιση, το ποσοστό των παρατηρήσεων που βρίσκονται μεταξύ 30 και 33; Τα δείγμα των τιμών της μεταβλητής είναι ομοιογενές;

14. Σε μία κανονική κατανομή είναι $\bar{x}=20$, $s=3$ Ποιο είναι, κατά προσέγγιση, το ποσοστό των παρατηρήσεων που βρίσκονται μεταξύ 14 και 26;

15. Σε μία κανονική κατανομή είναι $\bar{x}=25$, $s=5$ Ποιο είναι, κατά προσέγγιση, το ποσοστό των παρατηρήσεων που βρίσκονται μεταξύ 20 και 30;

16. Αν οι παρατηρήσεις μίας μεταβλητής Χ, σε ένα δείγμα μεγέθους ν, είναι $t_1, t_2, ..., t_\nu$, η μέση τιμή ισούται με:

(α) $\bar{x}=\dfrac{\sum_{i=1}^{\nu} t_i}{\nu}$ **(β)** $\bar{x}=\dfrac{\sum_{i=1}^{\nu} t_i}{\nu^2}$ **(γ)** $\bar{x}=\dfrac{\sum_{i=1}^{\nu} t_i^2}{\nu}$ **(δ)** $\bar{x}=\dfrac{\sum_{i=1}^{\nu} t_i^2}{\nu^2}$ **(ε)** $\bar{x}=\dfrac{\sum_{i=1}^{\nu} t_i}{\nu-1}$

17. Αν σε κάθε τιμή $x_1, x_2, \ldots, x_\nu$, ενός συνόλου δεδομένων, δοθεί διαφορετική βαρύτητα, εκφραζόμενη με τους συντελεστές βαρύτητας $w_1, w_2, \ldots, w_\nu$, ο σταθμικός μέσος δίνεται από τον τύπο:

(α) $\bar{x} = \dfrac{\sum_{i=1}^{\nu} x_i w_i}{\sum_{i=1}^{\nu} \nu_i}$ **(β)** $\bar{x} = \dfrac{\sum_{i=1}^{\nu} x_i w_i}{\nu}$ **(γ)** $\bar{x} = \dfrac{\sum_{i=1}^{\nu} x_i w_i}{\sum_{i=1}^{\nu} \nu_i^2}$ **(δ)** $\bar{x} = \dfrac{\sum_{i=1}^{\nu} x_i w_i}{\sum_{i=1}^{\nu} w_i}$

(ε) $\bar{x} = \dfrac{\sum_{i=1}^{\nu} x_i w_i}{\sum_{i=1}^{\nu} x_i}$

18. Πλήθος ν_1 παρατηρήσεων έχει μέση τιμή μ_1. Πλήθος ν_2 παρατηρήσεων έχει μέση τιμή μ_2. Πλήθος ν_3 παρατηρήσεων έχει μέση τιμή μ_3. Όλες οι παρατηρήσεις έχουν μέση τιμή μ. Αποδείξτε ότι $\mu_1 = \mu + \dfrac{\nu_2}{\nu_1}(\mu - \mu_2) + \dfrac{\nu_3}{\nu_1}(\mu - \mu_3)$.

19. Δίνονται οι τιμές $x_1, x_2, \ldots, x_\nu$, μίας μεταβλητής X, που έχουν μέση τιμή $\bar{x}$ και τυπική απόκλιση s. Αν οι τιμές $x_1^2, x_2^2, \ldots, x_\nu^2$ έχουν μέση τιμή $\overline{x^2}$, αποδείξτε ότι $s^2 = \overline{x^2} - \left(\bar{x}\right)^2$.

20. Δίνονται οι τιμές $x_1, x_2, \ldots, x_\nu$, μίας μεταβλητής X, που έχουν μέση τιμή $\bar{x}$ και τυπική απόκλιση s. Οι τιμές $x_1, x_2, \ldots, x_\nu$ έχουν, αντίστοιχα, σχετικές συχνότητες $f_1, f_2, \ldots, f_\nu$ αποδείξτε ότι $s^2 = \sum_{i=1}^{\nu} x_i^2 f_i - \left(\bar{x}\right)^2$.

21. Ποια είναι τα σπουδαιότερα μέτρα θέσεως;

22. Ποια είναι τα σπουδαιότερα μέτρα διασποράς (ή μεταβλητότητας);

23. Κατά την ομαδοποίηση παρατηρήσεων, αν είναι R το εύρος του δείγματος και k ο αριθμός των κλάσεων, το πλάτος των κλάσεων c είναι

(α) $c = \dfrac{k}{R}$, **(β)** $c = \dfrac{R}{k}$, **(γ)** $c = Rk$, **(δ)** $c = R + k$,

(ε) $c = R - k$ **(στ)** $c = k - R$

Προβλήματα στατιστικής.

1. Σπουδαστής διαθέτει τα χρήματα του ως εξής: Για ενοίκιο 20%, για φαγητό 30%, για ρούχα 30%, για διασκέδαση 15% και αποταμιεύει τα υπόλοιπα. Κάντε το αντίστοιχο κυκλικό διάγραμμα.

2. Ο χημικός τύπος του νιτρικού νατρίου είναι $NaNO_3$, δηλαδή το μόριο του αποτελείται από ένα άτομο νατρίου (Na), ένα άτομο αζώτου (N) και τρία άτομα οξυγόνου (O). Γράψτε κυκλικό διάγραμμα για τη σύσταση του $NaNO_3$.

3. Ο παρακάτω πίνακας δείχνει τη σχετική % συχνότητα της εβδομαδιαίας αποζημιώσεως 500 υπαλλήλων επιχειρήσεως. Συμπληρώστε τον πίνακα με τις συχνότητες (ν_i, f_i) και τις αθροιστικές συχνότητες $(N_i,\ \mathrm{F}_i,\ \mathrm{F}_i\%)$.

€/ week	$350 \le x < 360$	$360 \le x < 370$	$370 \le x < 380$	$380 \le x < 390$	$390 \le x < 400$
$f_i\%$	$f_1 = 8\%$	$f_2 = 28\%$	$f_3 = 44\%$	$f_4 = 16\%$	$f_5 = 4\%$

4. Η μέση ηλικία των 30 γυναικών που εργάζονται σε εταιρεία είναι 38 έτη, ενώ των 50 ανδρών 42 έτη. Ποια η μέση ηλικίας όλων των εργαζομένων στην εταιρεία;

5. Ο παρακάτω πίνακας δείχνει τις απουσίες 100 σπουδαστών. Αν η μέση τιμή των απουσιών των σπουδαστών είναι 34, υπολογίστε τα α, β.

Απουσίες	20	30	40	Σύνολο
Σπουδαστές	α	40	β	100

6. Αν πάρομε για παρατηρήσεις τους αριθμούς 3, 6, 6, 9, 2, 4, βρείτε την τυπική απόκλιση τους.

7. Ο αριθμός των παιδιών ενός δείγματος 100 οικογενειών φαίνεται στον παρακάτω πίνακα. Βρείτε τη μέση τιμή και την τυπική απόκλιση της κατανομής.

Παιδιά	0	1	2	3	4	5	6
Οικογένειες	18	20	30	15	10	5	2

8. Βρείτε τη μέση τιμή και την τυπική απόκλιση της κατανομής

Κλάσεις	$0 \le x < 10$	$10 \le x < 20$	$20 \le x < 30$	$30 \le x < 40$
Συχνότητες	4	3	2	1

9. Σε μία κατανομή η μέση τιμή ν παρατηρήσεων είναι α και η μέση τιμή μ παρατηρήσεων β. Αν οι παρατηρήσεις μ, ν αντιστοιχούν σε διαφορετικά άτομα, βρείτε τη μέση τιμή των $\mu+\nu$ παρατηρήσεων.

10. Σε ένα δείγμα μεγέθους ν η μέση τιμή είναι α. Αν για μ άτομα του δείγματος $(\mu < \nu)$ έχομε μέση τιμή β, να βρεθεί η μέση τιμή των υπολοίπων $\nu - \mu$ ατόμων του δείγματος.

11. Μία μεταβλητή x λαμβάνει τις ν το πλήθος τιμές $x_1, x_2, ..., x_\nu$. Αν α είναι το ελάχιστο και β το μέγιστο αυτών, δείξτε ότι $\alpha \le \bar{x} \le \beta$. Πότε ισχύει η ισότητα;

12. Αν η τυπική απόκλιση μίας μεταβλητής x είναι 0, δείξτε ότι η μέση τιμή ισούται με κάθε μία από τις τιμές της μεταβλητής.

13. Κατά τη διάρκεια μίας ημέρας καταγράψαμε τις εξής θερμοκρασίες σε ^{O}C 2, 5, 6, 6, 7, 5, 5, 4. Υπολογίστε τη μέση τιμή, τη διάμεσο, το εύρος, την τυπική απόκλιση και το συντελεστή μεταβολής.

14. Έστω x_1, x_2, x_3, x_4 οι τιμές της ποσοτικής μεταβλητής Χ που αφορά τα άτομα δείγματος με αντίστοιχες αθροιστικές συχνότητες $N_1 = 9,\ N_2 = 32,\ N_3 = 40,\ N_4 = 50$. Βρείτε το μέγεθος του δείγματος, τις συχνότητες και τις σχετικές συχνότητες των τιμών x_1, x_2, x_3, x_4.

15. Πέντε διαδοχικοί ακέραιοι αριθμοί έχουν μέση τιμή 5. Βρείτε τους αριθμούς αυτούς και τη διάμεσο τους. Αν σε κάθε έναν από τους δυο μεγαλύτερους από τους παραπάνω πέντε αριθμούς προσθέσομε τον αριθμό a, η μέση τιμή των πέντε αριθμών γίνεται 7. Δείξτε ότι $a = 5$.

16. Έστω δείγμα ν παρατηρήσεων κάθε μία από τις οποίες μπορεί να έχει τιμή 1, 3 ή 4. Βρείτε τη μέση τιμή $\overline{x}$ και το μέγεθος ν του δείγματος αν είναι γνωστό ότι αυτά είναι οι ρίζες της εξισώσεως $x^2 - 10x + 21 = 0$.

17. Η μέση βαθμολογία σπουδαστή σε 4 διαγωνίσματα μαθηματικών είναι 92 μονάδες.
Α. Αν στο πέμπτο διαγώνισμα γράψει 87, ποια η νέα μέση βαθμολογία;
Β. Πόσο πρέπει να γράψει στο πέμπτο διαγώνισμα για να ανεβάσει τη μέση βαθμολογία του κατά μία μονάδα;

18. Το μέσο βάρος 4 μαθητών είναι $60\ kg$. Έρχονται άλλοι δυο μαθητές που τα βάρη τους διαφέρουν κατά $2\ kg$ και το μέσο βάρος γίνεται $61\ kg$.
Α. Πόσο ζυγίζουν οι δυο νέοι μαθητές;

Β. Αν φύγει ένας μαθητής που ζυγίζει $59\ kg$, ποιο το μέσο βάρος των 5 μαθητών που απομένουν;

19. Το μέσο βάρος 6 κοριτσιών είναι $58\ kg$ και 3 αγοριών $61\ kg$. Βρείτε το συνολικό βάρος των κοριτσιών, το συνολικό βάρος των αγοριών και το μέσο βάρος αγοριών – κοριτσιών μαζί.

20. Το μέσο βάρος μία ομάδας ανθρώπων είναι $70\ kg$. Αν σε αυτούς προστεθεί ακόμη ένας που ζυγίζει $76\ kg$, το νέο βάρος είναι $71\ kg$. Πόσοι ήταν οι άνθρωποι της αρχικής ομάδας;

21. Ο μέσος μισθός 15 ανδρών και 10 γυναικών που απασχολούνται σε μία επιχείρηση είναι 900 €. Αν ο μέσος μισθός των ανδρών είναι 920 €, βρείτε το μέσο μισθό των γυναικών.

22. Η μέση τιμή πέντε αριθμών είναι 4 και η διάμεσος 3. Οι τρεις από αυτούς είναι οι αριθμοί 2, 5, 7. Βρείτε τους άλλους δυο. Αν στους πέντε παραπάνω αριθμούς προσθέσομε άλλους τρεις διαδοχικούς ακεραίους, η μέση τιμή των οκτώ αριθμών γίνεται 5,5. Ποιοι είναι οι αριθμοί που προσθέσαμε;

23. Επιχείρηση έχει προς ενοικίαση αυτοκίνητα για τα οποία ο μέσος χρόνος λειτουργίας τους πριν από την εμφάνιση της πρώτης βλάβης είναι 12 μήνες με τυπική απόκλιση 3 μήνες. Δείξτε ότι το δείγμα δεν είναι ομοιογενές.

24. Συμπληρώστε τους παρακάτω στατιστικούς πίνακες.

x_i	ν_i	N_i	f_i	F_i	$f_i\%$	F_i	$\nu_i x_i$
$x_1=5$	$\nu_1=$	$N_1=3$	$f_1=$	$F_1=$	$f_1\%=$	$F_1\%=$	$\nu_1x_1=$
$x_2=6$	$\nu_2=$	$N_2=10$	$f_2=$	$F_2=$	$f_2\%=$	$F_2\%=$	$\nu_2x_2=$
$x_3=7$	$\nu_3=$	$N_3=15$	$f_3=$	$F_3=$	$f_3\%=$	$F_3\%=$	$\nu_3x_3=$
$x_4=8$	$\nu_4=$	$N_4=20$	$f_4=$	$F_4=$	$f_4\%=$	$F_4\%=$	$\nu_4x_4=$
Σύνολο	$\nu=\sum_{i=1}^{4}\nu_i=$		$\sum_{i=1}^{4}f_i=1$		$\sum_{i=1}^{4}f_i\%=100$		$\sum_{i=1}^{4}\nu_i x_i=$

x_i	ν_i	N_i	f_i	F_i	$f_i\%$	F_i	$\nu_i x_i$
$x_1=5$	$\nu_1=$	$N_1=$	$f_1=0,1$	$F_1=$	$f_1\%=$	$F_1\%=$	$\nu_1x_1=$
$x_2=6$	$\nu_2=$	$N_2=$	$f_2=0,2$	$F_2=$	$f_2\%=$	$F_2\%=$	$\nu_2x_2=$
$x_3=7$	$\nu_3=$	$N_3=$	$f_3=$	$F_3=$	$f_3\%=$	$F_3\%=$	$\nu_3x_3=$
$x_4=8$	$\nu_4=$	$N_4=$	$f_4=0,3$	$F_4=$	$f_4\%=$	$F_4\%=$	$\nu_4x_4=$
Σύνολο	$\nu=\sum_{i=1}^{4}\nu_i=50$		$\sum_{i=1}^{4}f_i=1$		$\sum_{i=1}^{4}f_i\%=100$		$\sum_{i=1}^{4}\nu_i x_i=$

x_i	ν_i	N_i	f_i	F_i	$f_i\%$	F_i	$\nu_i x_i$
$x_1=1$	$\nu_1=10$	$N_1=$	$f_1=$	$F_1=$	$f_1\%=$	$F_1\%=$	$\nu_1x_1=$
$x_2=2$	$\nu_2=$	$N_2=35$	$f_2=$	$F_2=$	$f_2\%=$	$F_2\%=$	$\nu_2x_2=$
$x_3=3$	$\nu_3=$	$N_3=$	$f_3=$	$F_3=$	$f_3\%=$	$F_3\%=$	$\nu_3x_3=$
Σύνολο	$\nu=\sum_{i=1}^{3}\nu_i=50$		$\sum_{i=1}^{3}f_i=1$		$\sum_{i=1}^{3}f_i\%=100$		$\sum_{i=1}^{3}\nu_i x_i=$

x_i	ν_i	N_i	f_i	F_i	$f_i\%$	F_i	$\nu_i x_i$
$x_1=1$	$\nu_1=2$	$N_1=$	$f_1=0,1$	$F_1=$	$f_1\%=$	$F_1\%=$	$\nu_1x_1=$
$x_2=2$	$\nu_2=$	$N_2=8$	$f_2=$	$F_2=$	$f_2\%=$	$F_2\%=$	$\nu_2x_2=$
$x_3=3$	$\nu_3=8$	$N_3=$	$f_3=$	$F_3=$	$f_3\%=$	$F_3\%=$	$\nu_3x_3=$
$x_4=4$	$\nu_4=$	$N_4=$	$f_4=$	$F_4=$	$f_4\%=$	$F_4\%=$	$\nu_4x_4=$
Σύνολο	$\nu=\sum_{i=1}^{4}\nu_i=$		$\sum_{i=1}^{4}f_i=1$		$\sum_{i=1}^{4}f_i\%=100$		$\sum_{i=1}^{4}\nu_i x_i=$