Εισαγωγή

Οι κυνικοί λένε σαρκαστικά πως μπορείς να αποδείξεις οτιδήποτε με τη Στατιστική. Άλλοι πάλι υποστηρίζουν πως δεν μπορείς να κάνεις τίποτα με τη Στατιστική. Κάποιοι θυμίζουν ότι η Στατιστική είναι ένας τρόπος για να λέει κανείς ψέματα.

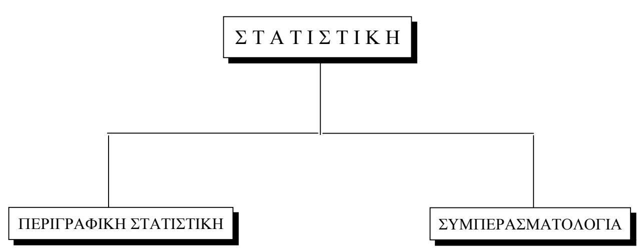
<u>Ένα παράδειγμα:</u> Η συντριπτική πλειοψηφία των ελλήνων έχει αριθμό κάτω άκρων μεγαλύτερο από τον μέσο όρο.
Πράγματι:
Υπάρχουν 500 συμπατριώτες μας που δεν έχουν κάτω άκρα (ακρωτηριασμοί κλπ). Δυόμισι χιλιάδες έχουν ένα πόδι. Οι υπόλοιποι (10 997 000) έχουν δύο πόδια.

Μέσος όρος κάτω άκρων των ελλήνων$= \dfrac{500\times 0+2500\times 1+10997000\times 2}{11000000} =1{,}999682.$

Η συντριπτική πλειοψηφία των ελλήνων έχει 2 πόδια, $2>1{,}999682$.

Αντικείμενο της Στατιστικής

Ας προχωρήσουμε όμως σε σοβαρότερα πράγματα Πρόκειται για κάποιες γνώσεις, γνωστές ίσως στους περισσότερους, που θα μας βοηθήσουν να τοποθετήσουμε τη στατιστική σε σωστά πλαίσια, έτσι ώστε να βοηθηθούν και οι μαθητές μας να καταλάβουν τη χρήση της αλλά και τη χρησιμότητά της.



Η **περιγραφική στατιστική** ασχολείται με τη συλλογή την ανάλυση και την ερμηνεία δεδομένων ενός συγκεκριμένου συνόλου. Η ανάλυση αυτή δεν γενικεύεται σε ευρύτερο σύνολο.

Η **συμπερασματολογία (επαγωγική στατιστική)** ασχολείται με την εξαγωγή συμπερασμάτων κάτω από συνθήκες αβεβαιότητας. Όταν ο στατιστικός παίρνει μια απόφαση δεν γνωρίζει την αλήθεια, γνωρίζει ποια από τις αποφάσεις που έχει μπροστά του είναι η ποιο πιθανή (παράδειγμα μια δημοσκόπηση).

Στη πλειονότητα των περιπτώσεων, ο σκοπός της συλλογής στοιχείων, αφορά κατά κύριο λόγο, **τη γενίκευση των συμπερασμάτων** για ένα ευρύτερο πληθυσμό, από κομμάτι του οποίου πήραμε τα στοιχεία. Όταν πάρουμε ένα **τυχαίο δείγμα**, και καταγράψουμε τις τιμές του μεγέθους που μας ενδιαφέρει (τις τιμές της τυχαίας μεταβλητής), ξέρουμε ότι δεν μπορούν να μας δώσουν, την πλήρη αλήθεια για τον πληθυσμό. Αυτό που θέλουμε είναι μια διαδικασία που θα μας λέει, σε ένα καθορισμένο επίπεδο εμπιστοσύνης, ότι η αλήθεια βρίσκεται μέσα σε κάποια συγκεκριμένα όρια.

Η επιλογή ενός τυχαίου δείγματος από ένα πληθυσμό για να μπορεί να δώσει πληροφορίες για τον πληθυσμό είναι μια δύσκολη και πολυσύνθετη υπόθεση.

Διαφοροποίηση των Μαθηματικών και της Στατιστικής

Η στατιστική δεν είναι ούτε θετική ούτε θεωρητική επιστήμη. Η στατιστική είναι στοχαστική επιστήμη. Μελετά τη μεταβλητότητα. Πειραματίζεται με πειράματα αβεβαίου αποτελέσματος. Κάτω από τις ίδιες συνθήκες είναι δυνατόν να καταλήξουν σε διαφορετικά αποτελέσματα.

Στα μαθηματικά συγκεκριμένες υποθέσεις οδηγούν σε μονοσήμαντο αποτέλεσμα.

Μέτρα κεντρικής θέσεως ή τάσεως

Η χρήση του «μέσου όρου» στην καθημερινή ζωή είναι αρκετά συχνή. “Το μέσο μηνιαίο εισόδημα ενός καθηγητή με 20 χρόνια υπηρεσίας είναι 1300€”, “ξοδεύω στο σουπερμάρκετ κατά μέσο όρο 550€ το μήνα”, “η μέση ανώτερη θερμοκρασία φέτος τον Ιανουάριο ήταν 9^0C”, “έβγαλα μέσον όρο 17,3 το πρώτο τετράμηνο”.

Στο πίσω μέρος του μυαλού, με τη χρήση του “μέσου όρου”, ενυπάρχει η έννοια ότι τα πράγματα τείνουν να συσσωρευτούν γύρω από μια κεντρική τιμή. Στη φυσική για παράδειγμα η έννοια αυτή έχει να κάνει με το κέντρο βάρους.

Στη στατιστική υπάρχουν διάφορα είδη “μέσων όρων”, διάφορα **μέτρα κεντρικής θέσεως**.

Α. Μέση τιμή (Αριθμητικός Μέσος)

Παράδειγμα:
Επτά δέματα ζυγίζουν 1,4,6,1,6,2,1 κιλά. Η μέση τιμή τους είναι 3 κιλά.

Αν προστεθεί ένα δέμα με βάρος 27κιλών ο αριθμητικός μέσος εκτινάσσεται στα 6 κιλά. Σκεφτείτε που θα φθάσει ο μέσος μισθός της περιοχής μας, αν ο πλουσιότερος άνθρωπο της χώρας έρθει να μείνει στο ίδια συνοικία με μας!

Η έννοια του αριθμητικού μέσου ενίοτε περιορίζεται όταν τα δεδομένα μας παίρνουν μόνον ακέραιες τιμές. Για παράδειγμα το δικτυακό τόπο του Λυκείου μας τον επισκέφθηκαν τον προηγούμενο μήνα, κατά μέσον όρο 6,6 άτομα ημερησίως (άρα περίπου επτά).

Τελικά η **μέση τιμή**:

α) Χρησιμοποιείται για όλες τις συνεχείς μεταβλητές και έχει περιορισμένη έννοια όταν πρόκειται για διακριτές μεταβλητές.

β) Για τον υπολογισμό της χρησιμοποιούνται όλες οι τιμές της μεταβλητής.

γ) Αρκετά μεγάλες ή αρκετά μικρές τιμές επηρεάζουν σημαντικά τη τιμή του αριθμητικού μέσου.

Υπάρχουν όμως και άλλοι **μέσοι**.

Παράδειγμα 1: Ταξίδεψε κάποιος από την Αθήνα στη Θεσσαλονίκη με μέση ταχύτητα 80km/h και επέστρεψε από την ίδια διαδρομή με μέση ταχύτητα 100km/h. Ποια ήταν η μέση ταχύτητα όλου του ταξιδιού;

Λύση: $\bar{\upsilon}=\frac{2s}{t_1+t_2}=\frac{2s}{\frac{s}{\upsilon_1}+\frac{s}{\upsilon_2}}=\frac{1}{\frac{1}{2}\left(\frac{1}{\upsilon_1}+\frac{1}{\upsilon_2}\right)}\simeq 88{,}89$ km/h (**Αρμονικός μέσος** των δύο ταχυτήτων).

Παράδειγμα 2: Τα κέρδη μιας εταιρείας σε τρία συνεχή έτη ήταν 20%, 25% και 45%. Ποιο ήταν το μέσο ετήσιο ποσοστό αύξησης αυτής της τριετίας;

Λύση: Αν το ζητούμενο ποσοστό είναι μ και α τα αρχικά κέρδη, τότε θα έχουμε την εξίσωση: $\alpha\left(1+\frac{20}{100}\right)\left(1+\frac{25}{100}\right)\left(1+\frac{45}{100}\right)=\alpha\left(1+\frac{\mu}{100}\right)^3$.

Επομένως $1+\frac{\mu}{100}=\sqrt[3]{\left(1+\frac{20}{100}\right)\left(1+\frac{25}{100}\right)\left(1+\frac{45}{100}\right)}$. (Ο **γεωμετρικός μέσος** των $1+\frac{20}{100}$, $1+\frac{25}{100}$ και $1+\frac{45}{100}$).

Παράδειγμα 3:

Στον διπλανό πίνακα δίνεται η εργασία 5 τεχνιτών τον μήνα Ιανουάριο, καθώς και η ωριαία αμοιβή τους:

Τεχνίτης	Ώρες εργασίας	Ωριαία αμοιβή (σε €)
Α	50	50
Β	150	50
Γ	200	40
Δ	50	80
Ε	50	80

Αν δεν ληφθούν υπ' όψιν οι ώρες εργασίας τον μήνα, η μέση ωριαία αμοιβή είναι $\frac{300}{5}=$ 60€. Αν λάβουμε όμως υπ' όψιν τις ώρες εργασίας κάθε εργάτη τότε υπολογίζουμε τον **σταθμισμένο αριθμητικό μέσο**: $\frac{50\times50+150\times50+200\times40+50\times80+50\times80}{50+150+200+50+50}=52€$.

Β. Διάμεσος

Ένα άλλο σπουδαίο μέτρο κεντρικής θέσεως είναι η **διάμεσος**. Είναι η μεσαία ή το ημιάθροισμα των δύο μεσαίων τιμών, ανάλογα αν είναι περιττό ή άρτιο αντίστοιχα το πλήθος των παρατηρήσεων. Με άλλα λόγια η διάμεσος κατέχει την κεντρική θέση, έτσι ώστε το 50% των τιμών να προηγούνται και το 50% των τιμών να ακολουθούν.

Στο προηγούμενο παράδειγμα (με τα επτά δέματα), η διάμεσος έχει τιμή 2. Αν προστεθεί η τιμή 27, η διάμεσος δεν αλλάζει αισθητά, γίνεται 3.

Τελικά η **διάμεσος:**

α) Χρησιμοποιείται για όλες τις συνεχείς μεταβλητές και για διακριτές μεταβλητές.

β) Όταν οι τιμές της μεταβλητής τοποθετηθούν κατά αύξουσα τάξη μεγέθους, η διάμεσος αγνοεί όλες τις παρατηρήσεις εκτός μίας ή δύο στο μέσο της σειράς των παρατηρήσεων

γ) Αρκετά μεγάλες ή αρκετά μικρές τιμές επηρεάζουν ελάχιστα την τιμή της διαμέσου.

δ) Η διάμεσος είναι το καλλίτερο μέτρο θέσεως για ασύμμετρες κατανομές (οικογενειακό εισόδημα, κατά κεφαλήν εισόδημα κλπ) ή όταν υπάρχουν τιμές σε μεγάλη απόσταση από τον κύριο όγκο των δεδομένων (έκτροπες).

Παράδειγμα από το Excel για τη σχέση μέσης τιμής και διαμέσου.

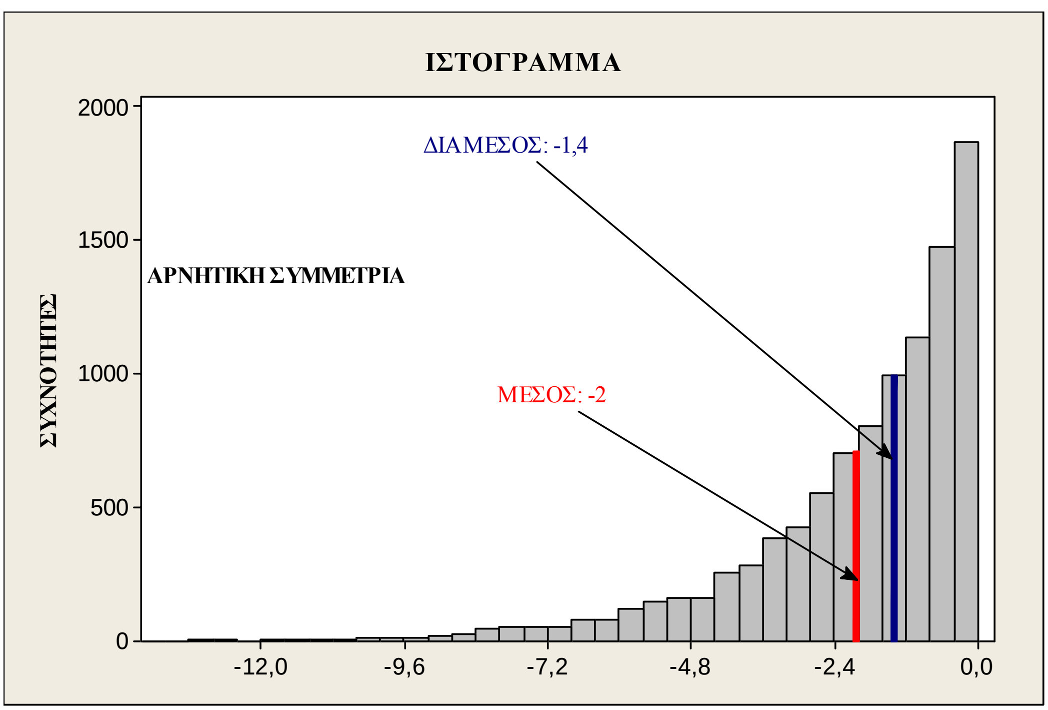
Σύγκριση Μέσου και Διαμέσου

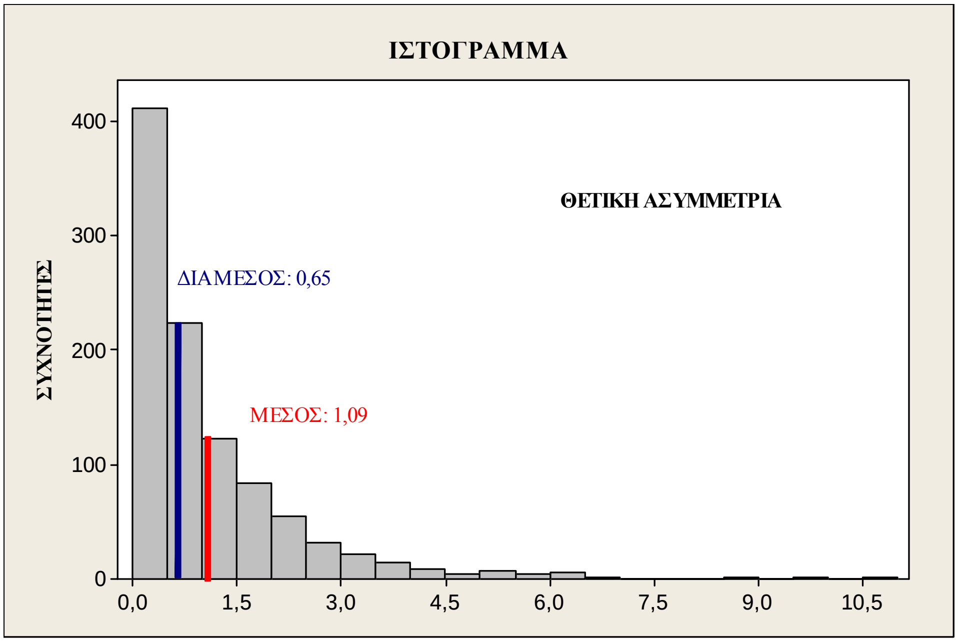
Μέσος	**Διάμεσος**
Ευαίσθητος στην επίδραση ακραίων τιμών, ειδικά σε μικρά σύνολα δεδομένων.	Όχι ευαίσθητη στην επίδραση ακραίων τιμών.
Λιγότερο αντιπροσωπευτικός ως «τυπική τιμή» για στρεβλές κατανομές με μία μόνο επικρατούσα τιμή.	Περισσότερο αντιπροσωπευτική ως «τυπική τιμή» για στρεβλές κατανομές με μια μόνο επικρατούσα τιμή.
Χρήσιμος για συμπερασματολογία που αναφέρεται στο άθροισμα των τιμών του πληθυσμού.	Όχι χρήσιμη για συμπερασματολογία που αναφέρεται στο άθροισμα των τιμών του πληθυσμού.
Ευκολότερος για να εργασθούμε με αυτόν θεωρητικά.	Δύσκολο να εργασθούμε με αυτήν θεωρητικά.

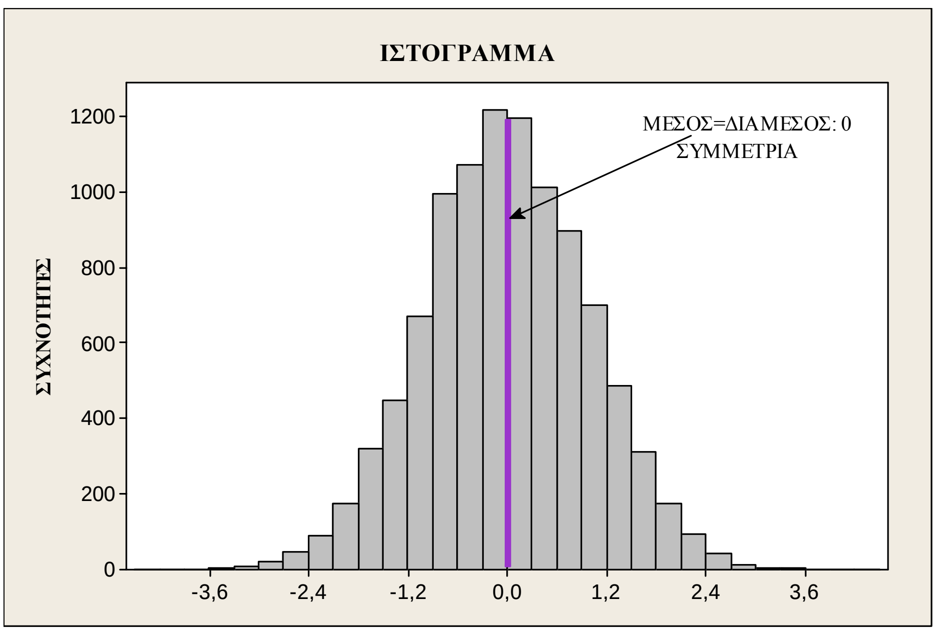
Μέτρα Ασυμμετρίας

3(Μέσος-Διάμεσος)/Τυπική Απόκλιση <u>Συντελεστής Pearson</u>
Θετικός: θετική ασυμμετρία κλπ

Σχέση μέσου, διαμέσου και συμμετρίας:







Μέτρα διασποράς

$$\textbf{Διακύμανση: } s^2 = \frac{1}{\nu}\sum_{i=1}^{\nu}\left(x_i - \bar{x}\right)^2$$

Χρησιμοποιείται όταν οι τιμές $x_1, x_2, \ldots, x_\nu$:
α) αποτελούν συνολικό πληθυσμό.
β) αποτελούν δείγμα από πληθυσμό και ενδιαφερόμαστε για τη διασπορά μέσα στο ίδιο το δείγμα.

Αν η διακύμανση του δείγματος $x_1, x_2, \ldots, x_\nu$ χρησιμοποιείται για την εκτίμηση της διακύμανσης του πληθυσμού, από τον οποίο προήλθε το δείγμα τότε χρησιμοποιείται ο τύπος:

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

ΠΡΟΣΟΧΗ:

α) Ο συντελεστής μεταβλητότητας είναι μέτρο **σχετικής διασποράς**.

β) $CV = \frac{s}{|\bar{x}|}$.

Κανονική Κατανομή

Η πρώτη ιστορικά εφαρμογή της κατανομής αυτής οφείλεται στον de Moivre (1733) οποίος διαπίστωσε ότι διωνυμικές πιθανότητες προσεγγίζονται ικανοποιητικά από την κανονική συνάρτηση πυκνότητας πιθανότητας. Μια δεύτερη ιστορικά σημαντική εφαρμογή της κανονικής κατανομής οφείλεται στον Gauss (1777-1855), ο οποίος διαπίστωσε ότι τα τυχαία σφάλματα στις μετρήσεις μιας ποσότητας ακολουθούν κανονική κατανομή. Για το λόγο αυτό η κανονική κατανομή είναι επίσης γνωστή ως Γκαουσιανή.

Η σημασία της κανονικής κατανομής

Η κανονική κατανομή είναι η σπουδαιότερη από όλες τις θεωρητικές κατανομές με τις περισσότερες εφαρμογές και αυτό οφείλεται κατά βάση στους εξής λόγους:

α) Οι εμπειρικές κατανομές ενός μεγάλου αριθμού μεταβλητών (οι οποίες αναφέρονται σε κοινωνικά, οικονομικά, φυσικά, βιολογικά κλπ φαινόμενα) περιγράφονται με ικανοποιητική προσέγγιση από τον κανονικό νόμου.

β) Υπό συγκεκριμένες προϋποθέσεις, η κανονική κατανομή αποτελεί ικανοποιητική προσέγγιση πολλών από τις γνωστές κατανομές (π.χ. διωνυμική, Poisson, κατανομή του student κλπ).

γ) Η κανονική κατανομή χρησιμοποιείται ευρύτατα στη συμπερασματολογία.

δ) Η κανονική κατανομή έχει ευρύτατη χρήση στις επιστήμες της υγείας.

Δυσκολίες κατανοήσεως εννοιών της Στατιστικής και συχνά λάθη

α) Από έρευνες που έχουν γίνει, φαίνεται να υπάρχει μία τάση των μαθητών να ανταποκρίνονται στα προβλήματα μαθηματικής φύσεως ανατρέχοντας στους υπολογισμούς με τη βοήθεια των τύπων, καθώς επίσης και σε μια ακολουθία διαδικασιών πριν ακόμα σχηματίσουν μια καθαρή εικόνα του προβλήματος. Είναι ικανοί να αποστηθίζουν τύπους και να ακολουθούν αλγοριθμικές διαδικασίες που ακολούθησαν σε παρόμοια προβλήματα με αυτά που έχουν διδαχτεί, αλλά σπάνια ανταποκρίνονται με επιτυχία σε προβλήματα στα οποία υπεισέρχονται οι καινούργιες καταστάσεις.

β) Υπάρχουν προβλήματα διατυπώσεως, καθώς και αδυναμίες κατανοήσεως του κειμένου.

γ) Δυσκολίες στις έννοιες των λόγων και των αναλογιών.

δ) Η διάμεσος.

-Δεν ταξινομούν τα δεδομένα κατά αύξουσα η φθίνουσα σειρά με αποτέλεσμα να δίνουν ως διάμεσο την τυχαία μεσαία παρατήρηση.

-Υπολογίζουν την διάμεσο σαν το ημιάθροισμα του πρώτου και του τελευταίου αριθμού στα ταξινομημένα δεδομένα.
-Δεν έχουν συνειδητοποιήσει ότι η διάμεσος δεν είναι απαραίτητα μια από τις τιμές των παρατηρήσεων.
-Έχουν δυσκολία στο να εντοπίζουν την μεσαία παρατήρηση η το ημιάθροισμα των μεσαίων παρατηρήσεων, ανάλογα αν το πλήθος είναι περιττό ή άρτιο.
-Συγχέουν την διάμεσο με το πλήθος των παρατηρήσεων με αποτέλεσμα να δίνουν σαν διάμεσο το μισό του πλήθους των παρατηρήσεων.

ε) Μέση Τιμή.

-Η μέση τιμή για τους μαθητές είναι περισσότερο μία υπολογιστική πράξη παρά μία εννοιολογική.
-Συνήθως η γνώση της μέσης τιμής εξαντλείται στην αποστήθιση του τύπου για τον υπολογισμό της.
-Στον πίνακα συχνοτήτων συγχέουν τις τιμές της μεταβλητής με τη συχνότητα εμφάνισης των τιμών.
-Προσθέτουν τις τιμές που αναφέρονται στις παρατηρήσεις και διαιρούν με το πλήθος τους, χωρίς να λάβουν υπόψη τους τις συχνότητες από τον πίνακα.
-Βρίσκουν το άθροισμα των τιμών των παρατηρήσεων και το διαιρούν με το άθροισμα των συχνοτήτων.
-Διαιρούν το άθροισμα των συχνοτήτων με το πλήθος τους.

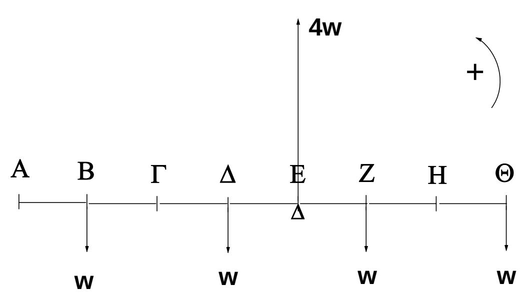
ς) Τα συνηθέστερα λάθη που γίνονται στις γραφικές παραστάσεις είναι:

-Σχήμα χωρίς τίτλο και πηγή πληροφοριών.
-Άξονες χωρίς αναφορά των μεταβλητών και χωρίς το μηδέν στην αρχή τους.
-Διαφοροποίηση του πλάτους στα ραβδογράμματα.

Ασκήσεις Στατιστικής

Η μέση τιμή στη Φυσική

Παράδειγμα :
Στην αβαρή ράβδο του σχήματος ΑΘ τα σημεία Β,Δ,Ζ και Θ απέχουν από την άκρη Α αποστάσεις 1cm, 3cm, 5cm και 7cm αντίστοιχα.



Σχ.1

1) Στις θέσεις Β,Δ,Ζ και Θ αναρτούμε ίσες μάζες, βάρους μέτρου w. Να βρεθεί το σημείο στο οποίο θα στηρίξουμε τη ράβδο ώστε να ισορροπεί.
2) Να βρεθεί η ακτίνα R ομογενούς δακτυλίου βάρους 4w, που έχει ροπή αδρανείας ως προς άξονα που διέρχεται από το κέντρο του και είναι κάθετος στο επίπεδο που ορίζει (Φυσική Γ΄ κατ. σελ118), ίση με την ροπή αδρανείας των παραπάνω μαζών ως προς άξονα κάθετο στη ράβδο που διέρχεται από το κέντρο βάρους της.
3) Στις θέσεις Β,Δ,Ζ και Θ τοποθετούμε τα βάρη w_1,w_2,w_3 και w_4 αντίστοιχα. Να βρεθεί το νέο κέντρο βάρους του συστήματος.
Λύση:

1) Αν x η απόσταση από την αρχή Α του σημείου στηρίξεως, τότε το αλγεβρικό άθροισμα των ροπών ως προς το Α θα είναι μηδέν:

Άρα $x4w-1w-3w-5w-7w=0 \Leftrightarrow x=\frac{1+3+5+7}{4}=4$ (η **μέση τιμή** των αποστάσεων).

2) Θα ισχύει $\frac{4w}{g}R^2=\frac{w}{g}(4-1)^2+\frac{w}{g}(4-3)^2+\frac{w}{g}(4-5)^2+\frac{w}{g}(4-7)^2 \Leftrightarrow$

$R=\sqrt{\frac{(4-1)^2+(4-3)^2+(4-5)^2+(4-7)^2}{4}}$ (η **τυπική απόκλιση** των αποστάσεων).

3) Αν x η απόσταση από την αρχή Α του σημείου στηρίξεως, τότε το αλγεβρικό άθροισμα των ροπών ως προς το Α θα είναι μηδέν:

Άρα $x(w_1+w_2+w_3+w_4)-1w_1-3w_2-5w_3-7w_4=0 \Leftrightarrow x=\frac{1w_1+3w_2+5w_3+7w_4}{w_1+w_2+w_3+w_4}$ (**ο σταθμισμένος αριθμητικός μέσος** των αποστάσεων με συντελεστές στάθμισης w_1, w_2, w_3 και w_4).

Άσκηση:

Να μελετηθεί η συνάρτηση $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ και να γίνει η γραφική της παράσταση, όπου $\sigma>0$ και $\mu\geq 0$.

Λύση:

α) Πεδίο ορισμού: **R**

β) $f(x+\mu)=f(x-\mu)$ για κάθε $x\in\mathbf{R}$, η ευθεία $x=\mu$ είναι άξονας συμμετρίας της C_f.

γ) $f'(x)=-\frac{x-\mu}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

δ) $f''(x)=\frac{1}{\sigma^3\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\left[\frac{(x-\mu)^2}{\sigma^2}-1\right]$.

ε) Η $y=0$ είναι οριζόντια ασύμπτωτη της C_f, όταν $x\to\pm\infty$.

ς) Ισχύει ο παρακάτω πίνακας μεταβολών:

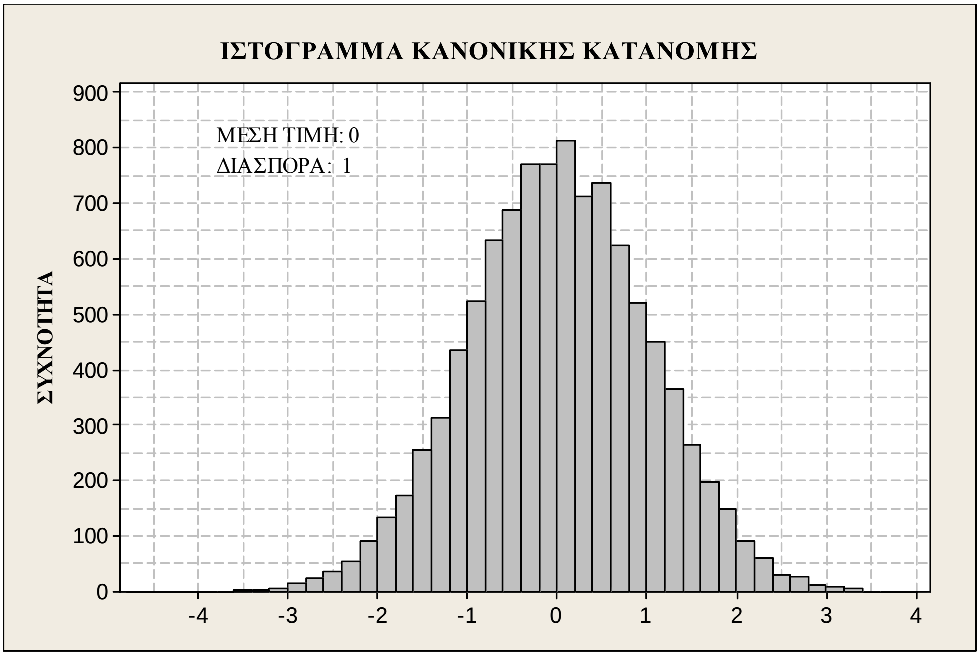
x	$-\infty$		$\mu-\sigma$		μ		$\mu+\sigma$		$+\infty$
$f'(x)$		+	|	+	ϕ	-	|	-	
$f''(x)$		+	ϕ	-	|	-	ϕ	+	
$f(x)$		$\nearrow$	|	$\nearrow$	|	$\searrow$	|	$\searrow$	
	0		Σ.Κ		max		Σ.Κ		0

$f(\mu)=\frac{1}{\sigma\sqrt{2\pi}}$, $f(\mu-\sigma)=\frac{1}{\sigma\sqrt{2\pi e}}=f(\mu+\sigma)$

ζ) <u>Γραφική Παράσταση για μ=0 και σ=1</u> (τυποποιημένη κανονική). Προτείνεται μήκος μονάδας στον x΄x τα 2cm και στον y΄y τα 8cm.

Να δοθούν $f(0) \simeq 0,4$, $f(1) \simeq 0,24$ και $f(1,5) \simeq 0,13$.
<u>Παρατήρηση:</u> Κομμάτι της ασκήσεως μπορεί να δοθεί και στα μαθηματικά γενικής παιδείας.

Εμβαδομέτρηση Ιστογράμματος κανονικής Κατανομής



Παρατηρήσεις σε δύο ασκήσεις του σχολικού βιβλίου:

α) Άσκηση 4, σελ. 103
Η άσκηση μπορεί να γενικευθεί για ν τιμές $x_1, x_2, \ldots, x_\nu$.

β) Άσκηση 7, σελ. 104
Όταν μετράμε το ύψος κάποιων ατόμων, λέμε 1 και 65 κλπ. Επίσης όταν μετράμε ηλικίες, λέμε 45 ετών ή 45 μισό. Οι μεταβλητές όμως «ύψος» και «ηλικία» είναι συνεχείς. Προκύπτει λοιπόν η ανάγκη να φτιάξουμε θεωρητικά όρια.
<u>Παράδειγμα</u>:
Να συμπληρωθεί ο πίνακας:

Πρακτικά όρια			Θεωρητικά όρια
x (ύψος σε cm)	ν_i	$f_i\%$	x (ύψος σε cm)
145-149	30		144,5-149,5
150-154	77		149,5-154,5
155-159	121		
160-164	135		κλπ
165-169	91		
170-174	30		
175-179	13		
180-184	3		
Σύνολο	500		