Τμήμα Αειφορικής Γεωργίας
Γεωπονική Σχολή
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

Εργαστηριακές Ασκήσεις Αναλυτικής και Οργανικής Χημείας

Αγγελική Απ. Γαλάνη
Χημικός PhD, Εργαστηριακό Διδακτικό Προσωπικό (ΕΔΙΠ)

7η Εργαστηριακή Άσκηση Στατιστική ανάλυση πειραματικών δεδομένων - Σφάλματα μετρήσεων

ΕΙΣΑΓΩΓΗ

Η πραγματική τιμή κάποιου μετρούμενου μεγέθους συνήθως δεν είναι γνωστή και αντί αυτής χρησιμοποιείται μια «παραδεκτή τιμή» με την οποία συγκρίνονται όλες οι πειραματικές τιμές

Μέση τιμή (mean) πειραματικών μετρήσεων $\bar{x} = \frac{x_1+x_2+\cdots+x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{\sum x_i}{N}$

Απόλυτο σφάλμα (error), E κάποιας μέτρησης x_i : $E_{xi} = x_i - \mu$

Σφάλμα μέσης τιμής $\bar{x}$: $E_{\bar{x}} = \bar{x} - \mu$

Ως ακρίβεια (accuracy), κάποιου αναλυτικού αποτελέσματος, ορίζεται η εγγύτητα της πειραματικής τιμής προς την αληθινή τιμή μ Εκφράζεται με το απόλυτο σφάλμα ή με αυτό της μέσης τιμής

Επαναληψιμότητα (precision), μετρήσεων. Δείχνει πόσο κοντά μεταξύ τους είναι τα αποτελέσματα. Μέτρα της επαναληψιμότητας μιας σειράς μετρήσεων, είναι κατά κύριο λόγο η μέση απόκλιση, η τυπική απόκλιση (standard deviation s) και το εύρος (range R).

Τυπική απόκλιση s, για μικρό αριθμό μετρήσεων

$$s = \pm\sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N-1}}$$

Όπου **N-1 οι βαθμοί ελευθερίας**

Για αριθμό μετρήσεων **N>20,** η μέση τιμή πλησιάζει την μ και η τυπική απόκλιση s την τιμή της **τυπικής απόκλισης του πληθυσμού σ,** η οποία δίνεται από τη σχέση

$$\sigma=\pm\sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}$$

Για μικρό αριθμό μετρήσεων, ένα χρήσιμο μέγεθος είναι και το **εύρος R (διαφορά ελάχιστης από τη μέγιστη τιμή)** $R = x_{μεγ} - x_{ελαχ}$

Τυπικό σφάλμα (Standard Error, SE) SE = s/√N

Σφάλμα, τυπική απόκλιση και εύρος → ίδιες μονάδες με το μέγεθος που μετριέται αλλά υπάρχουν φορές που είναι απαραίτητη η σύγκριση μεγεθών που εκφράζονται με διαφορετικές μονάδες, άρα είναι προτιμότερο να χρησιμοποιούνται οι σχετικές τιμές τους ως προς το μετρούμενο μέγεθος.

Έτσι προέκυψαν

- **<u>το σχετικό σφάλμα E_r</u>** $E_r = \frac{E}{\mu}$
- **<u>η σχετική τυπική απόκλιση S_r</u>** $s_r = \frac{s}{\bar{x}}$
- **<u>το σχετικό εύρος R_r</u>** $R_r = \frac{R}{\bar{x}}$

Αν τα παραπάνω σχετικά μεγέθη πολλαπλασιαστούν επί 100 ή 1000 ή 10^6 εκφράζονται επί τοις εκατό (%), επί τοις χιλίοις) ‰ και σε μέρη ανά εκατομμύριο

<u>Η % σχετική τυπική απόκλιση</u> $\frac{s}{\bar{x}}$ x100 με RSD % (% relative standard deviation) και είναι **<u>γνωστή ως συντελεστής μεταβλητότητας ή συντελεστής διασποράς CV (coefficient of variation)</u>**

<u>Χρήσιμη είναι και η διακύμανση που ορίζεται ως s^2</u>

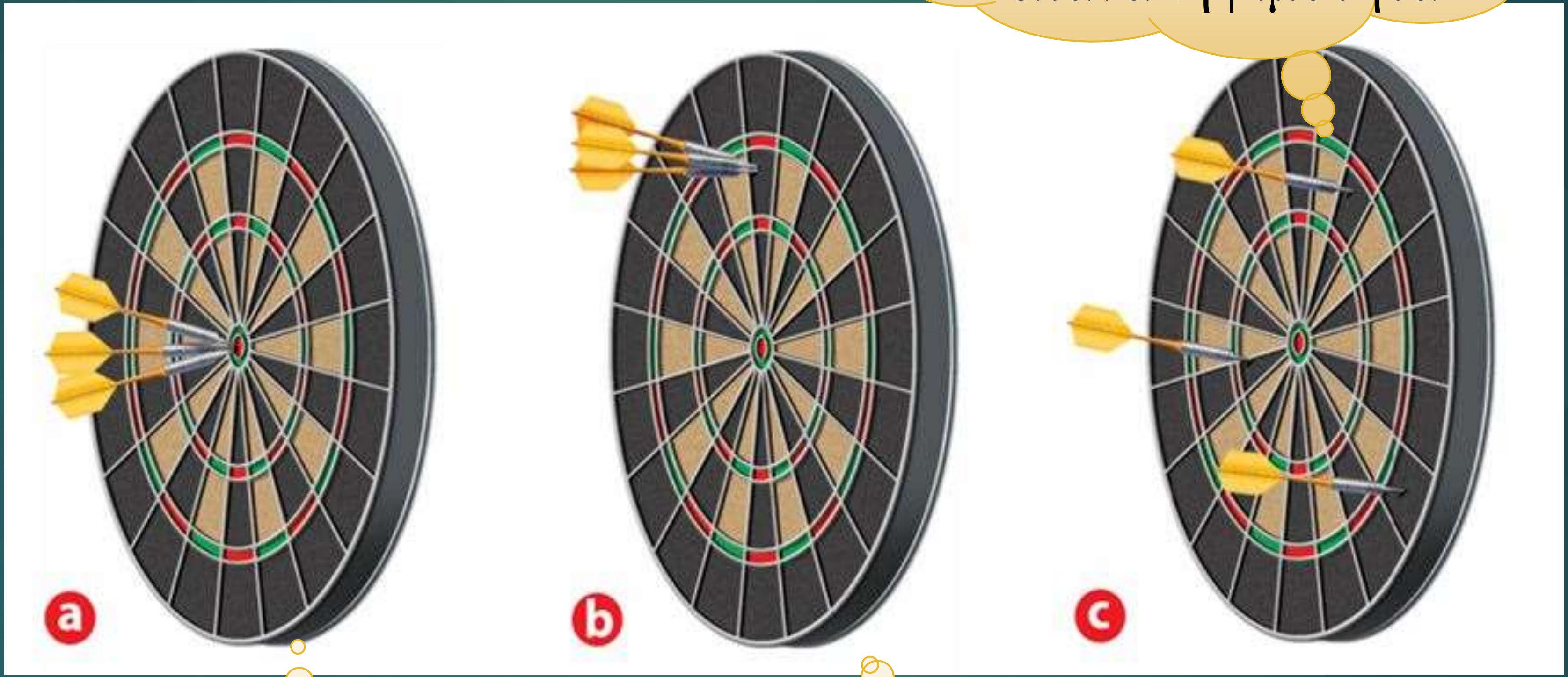
Προσοχή χρειάζεται στο **πρόσημο του αποτελέσματος στον υπολογισμό του σφάλματος μιας τιμής ή της μέσης τιμής**.

Όταν $\bar{x} < \mu$ το αποτέλεσμα έχει – πρόσημο και το σφάλμα είναι αρνητικό και όταν $\bar{x} > \mu$, το αποτέλεσμα έχει + πρόσημο και το σφάλμα είναι θετικό.

Ακρίβεια και επαναληψιμότητα είναι το μέτρο της αξιοπιστίας μιας αναλυτικής μεθόδου

- Ακρίβεια εκφράζεται συνήθως με το σφάλμα και με το σχετικό σφάλμα
- Επαναληψιμότητα εκφράζεται με το εύρος (για μικρό αριθμό μετρήσεων) και κατά κύριο λόγο με την τυπική απόκλιση και τη σχετική τυπική απόκλιση

Κακή ακρίβεια
Κακή
επαναληψιμότητα



Καλή ακρίβεια
Καλή
επαναληψιμότητα

Κακή ακρίβεια
Καλή
επαναληψιμότητα

Στατιστική Επεξεργασία Αποτελεσμάτων
Πίνακας τιμών μετρήσεων ογκομετρικής παροχής αντλίας
Δεδομένα προερχόμενα από: Νταρακάς Ευθύμιος, Πεταλά Μαρία, Τσιρίδης Βασίλειος, «Περιβαλλοντική Χημεία και Μηχανική», 2020, Εκδόσεις Τζόλα

Μέτρηση	x_i	x_i - x	$(x_i – x)^2$
1η	15.11	-0.01	0.0001
2η	15.15	0.03	0.0009
3η	15.12	0.00	0.0000
4η	15.13	0.01	0.0001
5η	15.09	-0.03	0.0009
Σύνολο	75.59	0	0.0020
x=75.59 /5 = 15.12 s= 0.002/4=0.022 mL/min			

- Σε πολλές περιπτώσεις ο μέσος όρος εκφράζεται με την αριθμητική του τιμή ± τυπική απόκλιση. Έτσι φαίνεται το διάστημα διασποράς των τιμών από το μέσο όρο.
- Στο παράδειγμα του διπλανού Πίνακα το αποτέλεσμα της μέτρησης της ογκομετρικής παροχής της αντλίας είναι:

  15.12 ± 0.022 mL/min.

Εκτός της έκφρασης του μέσου όρου με την αριθμητική του τιμή ± την τυπική απόκλιση, μπορεί να οριστεί κι άλλο διάστημα τιμών αυξημένης πιθανότητας να περιλαμβάνεται η πραγματική τιμή της μετρούμενης παραμέτρου.

- Είναι δυνατόν να οριστεί η πιθανότητα ακρίβειας (για παράδειγμα 95% ή 99% ή και η πιθανότητα σφάλματος 5% ή 1%). Το διάστημα των τιμών στο οποίο περιέχεται η πραγματική τιμή, ονομάζεται % διάστημα εμπιστοσύνης (Confidence Interval, Cl). Συνηθίζεται η επιλογή του 95% ή του 99%.
- Όταν ο αριθμός δειγμάτων για μια ανάλυση είναι μεγάλος και τα αποτελέσματα ακολουθούν κατανομή κανονική είναι αρκετά ακριβές να θεωρηθεί πως το 95% των μετρήσεων είναι μεταξύ x ± 2.SE και το 99% των μετρήσεων είναι μεταξύ x ± 3.SE
  - Όπου SE είναι το τυπικό σφάλμα (Standard Error) SE= $s/\sqrt{n}$

- Όταν το δείγμα ακολουθεί κανονική κατανομή, το διάστημα εμπιστοσύνης δίνεται από τη σχέση: %CI = x ± $t_{\alpha(2),(n-1)}$. SE,

Όπου: $t_{\alpha(2),(n-1)}$: τιμή που την παίρνουμε από στατιστικούς πίνακες και εξαρτάται από την πιθανότητα σφάλματος (π.χ. 5% ή 1%) και τον αριθμό των μετρήσεων, (n-1).

  - Για κανονική κατανομή (δηλαδή σε πολύ μεγάλο αριθμό μετρήσεων, πάνω από 30), ο όρος $t_{\alpha(2),(n-1)}$ ισούται με 1.96 για πιθανότητα σφάλματος 5% και με 2.576 για πιθανότητα σφάλματος 1%.
  - Για λιγότερες μετρήσεις (για παράδειγμα για 10), ο όρος $t_{\alpha(2),(n-1)}$ ισούται με 2,262 για πιθανότητα σφάλματος 5% και με 3.250 για πιθανότητα σφάλματος 1%

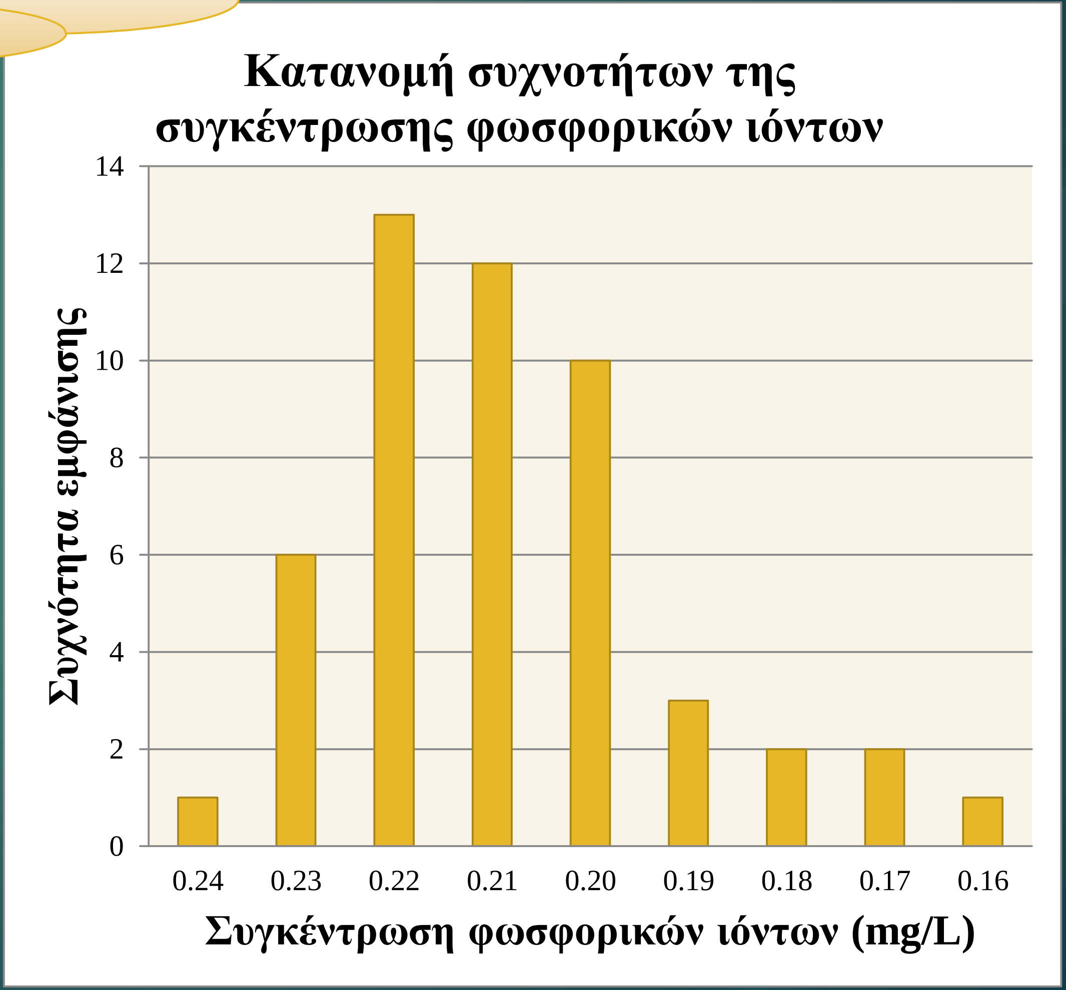
Κανονική (Gauss)κατανομή πληθυσμού

Σε δείγμα νερού λίμνης ελήφθησαν 50 αποτελέσματα μέτρησης συγκέντρωσης φωσφορικών ιόντων (προέκυψαν μετά από επαναληπτικές μετρήσεις

Παράδειγμα από: Νταρακάς Ευθύμιος, Πεταλά Μαρία, Τσιρίδης Βασίλειος, «Περιβαλλοντική Χημεία και Μηχανική», 2020, Εκδόσεις Τζόλα)

Μέσος όρος μετρήσεων $\bar{x}$ = 0,21 mg/L
Τυπική απόκλιση s = ± 0.017 mg/L

Συγκέντρωση Φωσφορικών ιόντων (mg/L)	Συχνότητα εμφάνισης αποτελέσματος
0.24	1
0.23	6
0.22	13
0.21	12
0.20	10
0.19	3
0.18	2
0.17	2
0.16	1



- Πληθυσμός, ονομάζεται το σύνολο των μετρήσεων
- Δείγμα, ονομάζονται οι 50 μετρήσεις που πραγματοποιήθηκαν
- Ο μέσος όρος των αποτελεσμάτων αντιστοιχεί στην πραγματική τιμή των φωσφορικών ιόντων στο νερό της λίμνης (με την προϋπόθεση ότι υπάρχουν μόνο τυχαία σφάλματα).

* X η όποια τιμή δείγματος έχει τη μορφή του σχήματος της προηγούμενης διαφάνειας
* n ο αριθμός μετρήσεων

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

Στο παράδειγμα εκτίμηση του μέσου όρου του πληθυσμού μ, είναι δυνατόν να αποτελεί ο υπολογισμός του μέσου όρου των δειγμάτων x

- Η τυπική απόκλιση σ, μπορεί να υπολογιστεί από την εξίσωση του διπλανού σχήματος.

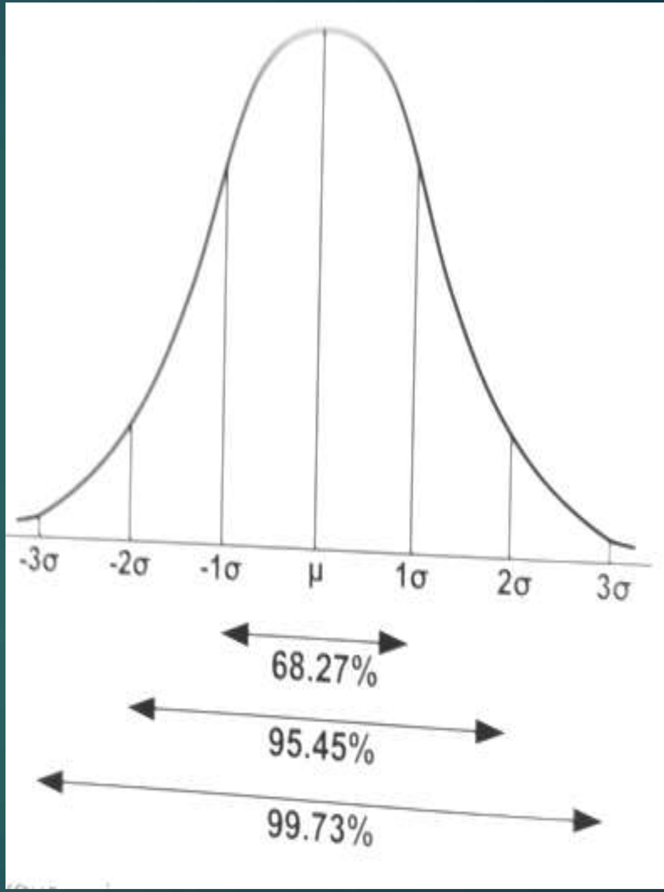
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Η μορφή κατανομής των αποτελεσμάτων υπολογίζεται με τη βοήθεια του ιστογράμματος της προηγούμενης διαφάνειας που είναι αντιπροσωπευτικό κανονικής κατανομής κατά Gauss

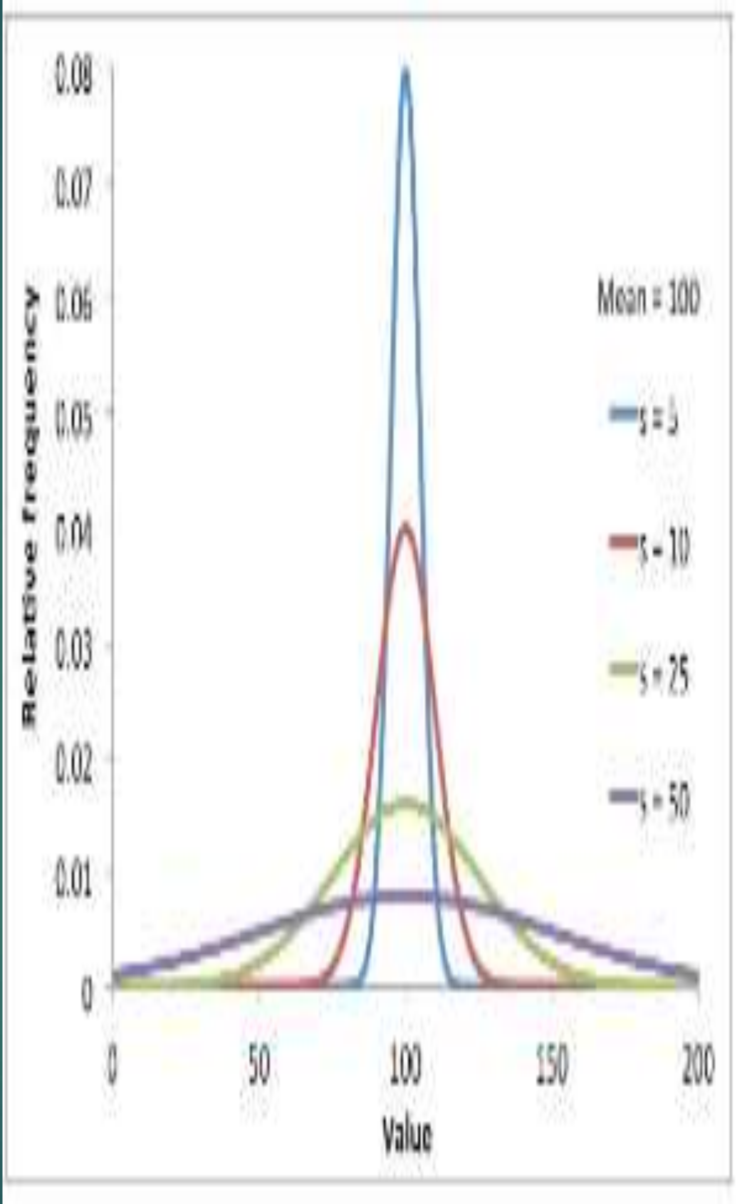
- Η κανονική κατανομή περιγράφεται από τη συνάρτηση του διπλανού σχήματος

Γενικές ιδιότητες κανονικής κατανομής

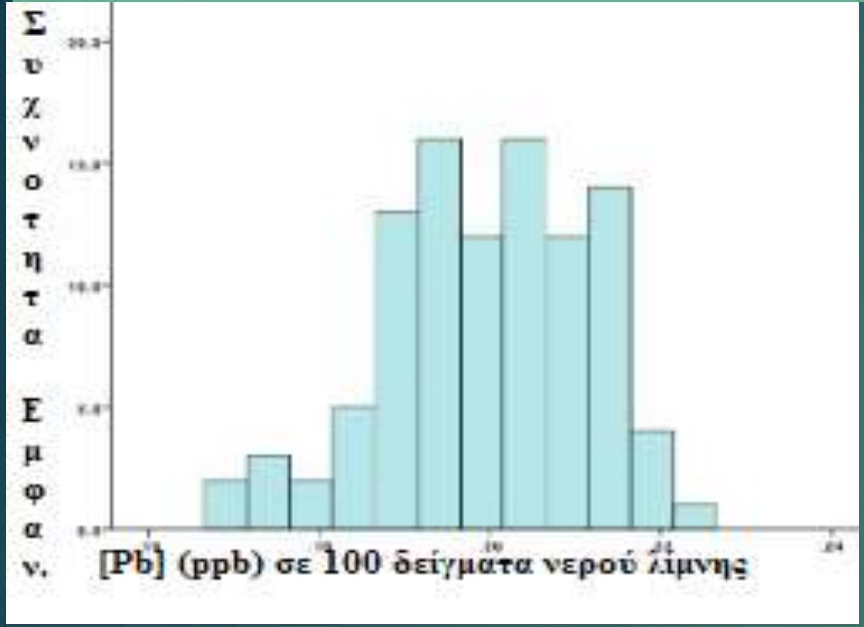
- Η καμπύλη επικεντρώνεται γύρω από μια κορυφή ίση με τη μέση τιμή και έχει χαρακτηριστικό κωδωνοειδές σχήμα
- Η καμπύλη είναι ομαλή και δεν έχει όρια στον άξονα χ
- Η πλειονότητα των εμφανιζόμενων μετρήσεων συγκεντρώνεται εντός τριών τυπικών αποκλίσεων του μέσου όρου
- Το μέγεθος της συχνότητας μειώνεται απότομα στο διάστημα μ ± 1σ και σε μεγαλύτερα διαστήματα ομαλότερα
- Σχεδόν το 68% του πληθυσμού , είναι σε απόσταση ± 1σ από το μέσο όρο μ
- Σχεδόν το 95% του πληθυσμού , είναι σε απόσταση ± 2σ από το μέσο όρο μ
- Σχεδόν το 99,7% του πληθυσμού , είναι σε απόσταση ± 3σ από το μέσο όρο μ

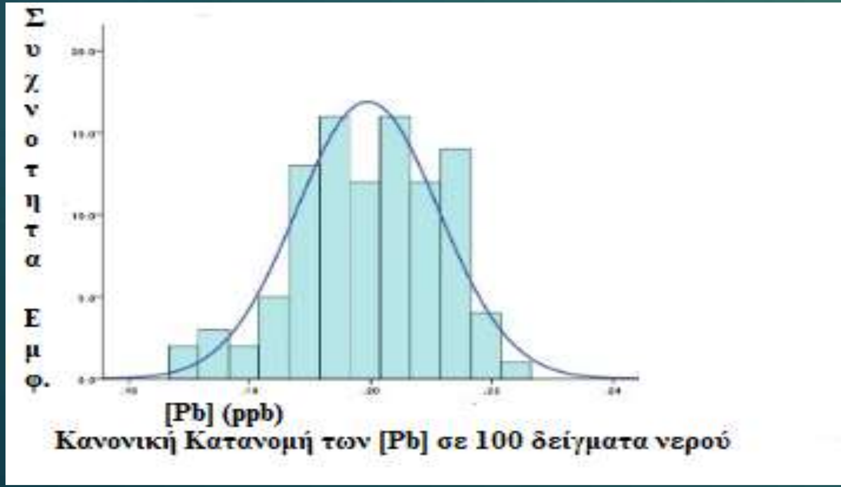


- Η καμπύλη κανονικής κατανομής, είναι δυνατόν να πάρει μορφές που να διαφέρουν ως προς το ύψος και το πλάτος.
- Το πλάτος εξαρτάται από το μέγεθος της τυπικής απόκλισης. Στο διπλανό σχήμα φαίνεται πώς όσο αυξάνει η τυπική απόκλιση η καμπύλη γίνεται ευρύτερη και μικρότερη (σε ύψος). Μεγαλύτερη τυπική απόκλιση υποδηλώνει ύπαρξη περισσότερων μετρήσεων που βρίσκονται σχετικά μακριά από τη μέση τιμή. Μικρότερη τυπική απόκλιση χαρακτηρίζει σύνολο δεδομένων με τις μετρήσεις πιο κοντά στον μέσο όρο.
- Για να μπορούν να συγκριθούν οι καμπύλες τα αποτελέσματα κανονικοποιούνται με βάση τον μετασχηματισμό: $z = (X-\mu)/\sigma$. Όπου z είναι το κανονικοποιημένο αποτέλεσμα.



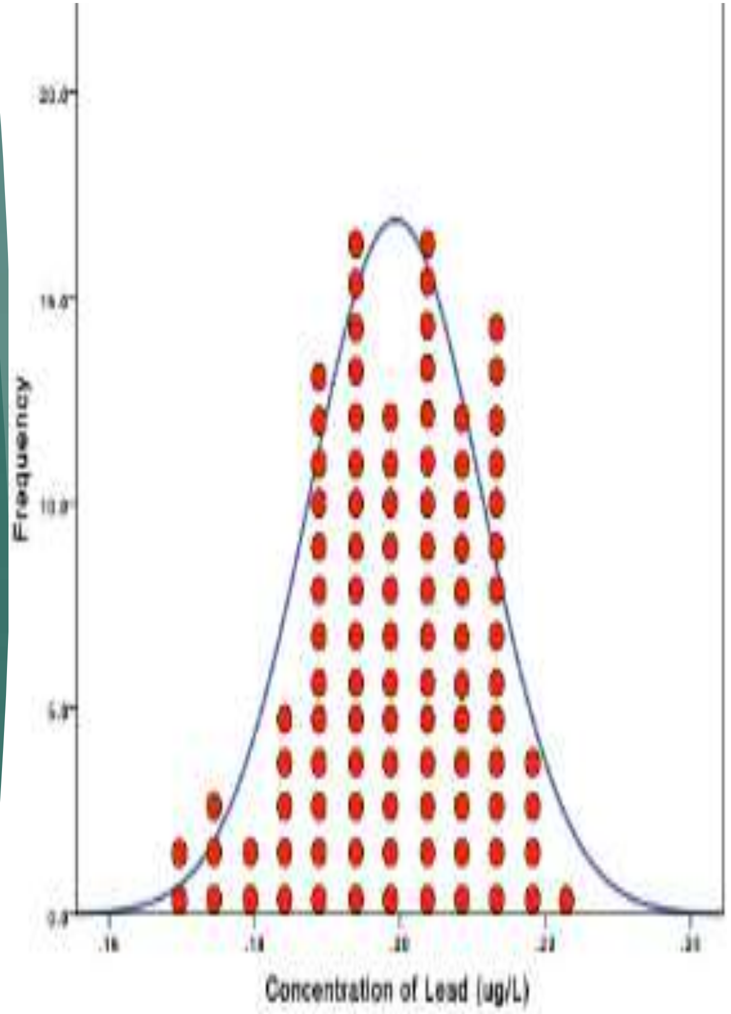
Τα παρακάτω σχήματα, αφορούν σε μετρήσεις συγκέντρωσης μολύβδου σε 100 δείγματα νερού.





- Το πρώτο είναι το ιστόγραμμα των τιμών, δηλαδή απεικονίζει τη συχνότητα εμφάνισης (1φορά, 2, 3,…κλπ) κάποιας συγκεκριμένης τιμής [Pb]
- Στο $2^{ο}$ φαίνεται και η καμπύλη κανονικής κατανομής των τιμών [Pb]
- Το δείγμα εδώ είναι οι 100 μετρήσεις
- Ο μέσος όρος των αποτελεσμάτων αντιστοιχεί στην πραγματική τιμή του μόλυβδου στο νερό της λίμνης με την προϋπόθεση πως υπήρχαν μόνο τυχαία σφάλματα.

- Μια κανονική κατανομή μπορεί επίσης να θεωρηθεί ως συνάρτηση πυκνότητας. Η περιοχή κάτω από την καμπύλη είναι σταθερή για όλες τις κανονικές κατανομές καθώς είναι πάντα ίση με 1 (εξ ορισμού). Ένα βασικό χαρακτηριστικό μιας κανονικής κατανομής είναι ότι η περιοχή που ορίζεται από δύο τιμές κατά μήκος του άξονα x αντιπροσωπεύει την πιθανότητα μια τυχαία κατανεμημένη μέτρηση να εμπίπτει σε αυτό το εύρος.



Όρια εμπιστοσύνης με την προϋπόθεση πως το δείγμα είναι αρκετά μεγάλο ώστε η πραγματική τυπική απόκλιση σ να ισούται με την τυπική απόκλιση του δείγματος

Όρια εμπιστοσύνης = $\bar{x} \pm z.(\sigma/\sqrt{n})$

όπου:

- ✓ $z = 1.96$ για 95% όρια εμπιστοσύνης
- ✓ $z = 2.58$ για 99% όρια εμπιστοσύνης
- ✓ $z = 2.97$ για 99.7% όρια εμπιστοσύνης

Όρια εμπιστοσύνης εάν το δείγμα δεν είναι αρκετά μεγάλο και εάν εκτιμάται πως η πραγματική τυπική απόκλιση σ απέχει από την τυπική απόκλιση του δείγματος

Όρια εμπιστοσύνης =

$$= \bar{x} \pm t_{n-1} \cdot (s/\sqrt{n})$$

Βαθμοί ελευθερίας (n-1)	Τιμή συντελεστή t για 95% όρια εμπιστοσύνης	Τιμή συντελεστή t για 99% όρια εμπιστοσύνης
2	4.30	9.92
5	2.57	4.03
10	2.23	3.17
20	2.09	2.85
50	2.01	2.68
100	1.98	2.63

Κανόνες γραφής σημαντικών ψηφίων κατά την παρουσίαση αποτελεσμάτων

Δεκαδικοί αριθμοί:

Αναφέρονται μόνο τα σημαντικά ψηφία του αριθμού

- Τερματικά μηδενικά δεξιά της υποδιαστολής είναι πάντα σημαντικά
- Σε περίπτωση που το μηδέν βρίσκεται ανάμεσα σε μη μηδενικά ψηφία, λογίζεται ως σημαντικό ψηφίο
- Σε περίπτωση που το μηδέν βρίσκεται αριστερά κάποιου δεκαδικού αριθμού, δεν λογίζεται ως σημαντικό ψηφίο

π.χ. 5,11 cm 5,00 cm, 0,101 cm 0,000412 cm όλοι 3 σημαντικά ψηφία

ΚΑΝΟΝΕΣ ΚΑΘΟΡΙΣΜΟΥ ΣΗΜΑΝΤΙΚΩΝ ΨΗΦΙΩΝ

Μη δεκαδικοί αριθμοί:

- Τερματικά μηδενικά σε μη δεκαδικούς αριθμούς, μπορεί να είναι σημαντικά μπορεί και όχι.

  Για παράδειγμα έστω μετράμε την ωμική αντίσταση R = 21000 Ohm. Δεν γνωρίζουμε αν τα τρία τελευταία μηδενικά είναι σημαντικά ψηφία. Το αποτέλεσμα θα πρέπει να γραφεί ως δύναμη του 10

  - ✓ 2.1×10^4 ο αριθμός έχει δύο σημαντικά ψηφία και βέβαιο είναι μόνο το πρώτο
  - ✓ 2.10×10^4 ο αριθμός έχει τρία σημαντικά ψηφία και αβέβαιο είναι το τελευταίο μηδενικό

Σημαντικά ψηφία σε υπολογισμούς

- Σε πολλαπλασιασμό ή διαίρεση, το αποτέλεσμα δίνεται με τόσα σημαντικά ψηφία όσα έχει η μέτρηση με τα λιγότερα σημαντικά.

- Σε πρόσθεση ή αφαίρεση το αποτέλεσμα δίνεται με τόσα σημαντικά ψηφία όσα έχει η μέτρηση με τα λιγότερα δεκαδικά ψηφία

Στρογγυλοποίηση

- Έτσι ονομάζεται η διαδικασία απόρριψης ψηφίων που δεν είναι σημαντικά από το αποτέλεσμα υπολογισμών και η τροποποίηση του τελευταίου ψηφίου που μένει
- Όλη η προσοχή δίνεται στο πρώτο από αριστερά ψηφίο που πρέπει να απορρίψουμε.
- Εάν αυτό είναι 5 ή μεγαλύτερο, προσθέτουμε μια μονάδα στο ψηφίο που προηγείται και και απορρίπτουμε όσα ψηφία βρίσκονται δεξιά αυτού.

Στρογγυλοποίηση σε τρία σημαντικά του 3,3453 δίνει 3,35

- Εάν είναι μικρότερο του 5 απλώς το απορρίπτουμε με όλα τα υπόλοιπα που βρίσκονται δεξιά του

Στρογγυλοποίηση του 3,3443 σε τρία σημαντικά ψηφία δίνει 3,34

ΠΕΙΡΑΜΑΤΑΙΚΟ ΜΕΡΟΣ

- Να πραγματοποιήσετε 5 φορές τη ζύγιση (σε ζυγό 4 δεκαδικών ψηφίων), συγκεκριμένου αντικειμένου το οποίο θα σας υποδείξουν στο Εργαστήριο.

Λαμβάνοντας υπόψιν το πραγματικό βάρος του αντικειμένου το οποίο θα σας γίνει γνωστό, να υπολογίσετε τη μέση τιμή, το εύρος, το σχετικό εύρος (%), την τυπική απόκλιση, τη σχετική τυπική απόκλιση (%), το σφάλμα της μέσης τιμής και το σχετικό (%) σφάλμα της.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Καλοκαιρινός Αντώνιος, Καθηγητής Αναλυτικής Χημείας ΕΚΠΑ « Αναλυτική Χημεία», Ανοικτές Ακαδημαϊκές Εκδόσεις Κάλλιπος, https://repository.kallipos.gr/bitstream/11419/6283/3/13044_Kalokairinos-final-KOY.pdf
- Νταρακάς Ευθύμιος, Πεταλά Μαρία, Τσιρίδης Βασίλειος, «Περιβαλλοντική Χημεία και Μηχανική», 2020, Εκδόσεις Τζόλα
- Καθ. Κωνσταντίνου Η. Ευσταθίου, Εργαστήριο Αναλυτικής Χημείας Πανεπιστημίου Αθηνών «ΕΝΟΤΗΤΑ Α ΣΤΑΤΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΑΝΑΛΥΤΙΚΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ», Ανοικτά Ακαδημαϊκ΄ά Μαθήματα ΕΚΠΑ
- https://sites.chem.utoronto.ca/chemistry/coursenotes/analsci/stats/MeasMeanVar.html
- https://groups.chem.ubc.ca/stoodley/Module1_3_Gaussian.html
- https://www.expii.com/t/types-of-error-overview-comparison-8112
- «Εργαστηριακές Ασκήσεις Φυσικής», Τόμος Ι, Εθνικό Μετσόβειο Πολυτεχνείο, Τομέας Φυσικής Γενικό Τμήμα, Εκδόσεις Συμμετρία, Αθήνα 1994
- «Ανάλυση Σφαλμάτων», Ε-Π Χριστοπούλου, ΄Καθηγήτρια Τμήματος Φυσικής Σχολής Θετικών επιστημών Πανεπιστημίου Πατρών, Μάρτιος 2005