

Τμήμα Επιστήμης Τροφίμων & Διατροφής

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΑΝΑΛΥΤΙΚΗ ΧΗΜΕΙΑ

Ενότητα 2: Στατιστική – Επεξεργασία δεδομένων

Δημήτρης Π. Μακρής *PhD DIC*

Αναπληρωτής Καθηγητής



© 2019 - 2020

1. Θεμελιώδεις μονάδες μέτρησης

Οι μετρήσεις συνήθως αποτελούνται από την μονάδα και έναν αριθμό που εκφράζει την ποσότητα αυτής της μονάδας. Υπάρχουν διάφορες μονάδες για την έκφραση μιας μέτρησης.

Για λόγους συνέπειας και αποφυγής σύγχυσης, υπάρχει μια κοινή ομάδα θεμελιωδών μονάδων που ονομάζονται μονάδες SI (SI units). Όλες οι μονάδες εκφράζονται με βάση τις θεμελιώδεις.

SI Base Units		
Physical Quantity	Name of Unit	Abbreviation
Mass	kilogram	kg
Length	meter	m
Time	second	s
Temperature	kelvin	K
Amount of substance	mole	mol
Electric current	ampere	A
Luminous intensity	candela	cd

SI-derived units with special names				
Quantity	Unit	Symbol	Expression in terms of other units	Expression in terms of SI base units
Frequency	hertz	Hz		1/s
Force	newton	N		m · kg/s ²
Pressure	pascal	Pa	N/m ²	kg/(m · s ²)
Energy, work, quantity of heat	joule	J	N · m	m ² · kg/s ²
Power, radiant flux	watt	W	J/s	m ² · kg/s ³
Quantity of electricity, electric charge	coulomb	C		s · A
Electric potential, potential difference, electromotive force	volt	V	W/A	m ² · kg/(s ³ · A)
Electric resistance	ohm	Ω	V/A	m ² · kg/(s ³ · A ²)
Electric capacitance	farad	F	C/V	s ⁴ · A ² /(m ² · kg)

1. Θεμελιώδεις μονάδες μέτρησης

Στην Χημεία συχνά χρησιμοποιούνται μετρήσεις που είναι είτε πολύ μεγάλες είτε πολύ μικρές. Για λόγους απλότητας, αυτές οι μετρήσεις εκφράζονται με την λεγόμενη επιστημονική σημειογραφία (**scientific notation**).

Επίσης, είναι δυνατόν να εκφραστούν οι μετρήσεις με την χρήση προθεμάτων.

Prefixes					
Prefix	Symbol	Factor	Prefix	Symbol	Factor
yotta	Y	10^{24}	deci	d	10^{-1}
zetta	Z	10^{21}	centi	c	10^{-2}
exa	E	10^{18}	milli	m	10^{-3}
peta	P	10^{15}	micro	μ	10^{-6}
tera	T	10^{12}	nano	n	10^{-9}
giga	G	10^9	pico	p	10^{-12}
mega	M	10^6	femto	f	10^{-15}
kilo	k	10^3	atto	a	10^{-18}
hecto	h	10^2	zepto	z	10^{-21}
deca	da	10^1	yocto	y	10^{-24}

2. Σημαντικά ψηφία

Η καταγραφή μιας μέτρησης δίνει πληροφορίες για amφότερα το μέγεθος και την αβεβαιότητά της. Για παράδειγμα, εάν ζυγίσουμε ένα δείγμα σε ένα ζυγό και η μάζα του είναι 1.2637 g, υποθέτουμε ότι όλα τα ψηφία, εκτός του τελευταίου είναι ακριβώς γνωστά. Υποθέτουμε ότι το τελευταίο ψηφίο έχει μια αβεβαιότητα τουλάχιστον ± 1 , δίνοντας μια απόλυτη αβεβαιότητα τουλάχιστον ± 0.0001 g, ή αλλιώς μια σχετική αβεβαιότητα τουλάχιστον:

$$\frac{\pm 0.0001 \text{ g}}{1.2637 \text{ g}} \times 100 = \pm 0.0079\%$$

Τα **σημαντικά ψηφία** είναι η αντανάκλαση της αβεβαιότητας μιας μέτρησης. Ο αριθμός των σημαντικών ψηφίων ισούται με τον αριθμό των ψηφίων της μέτρησης, με την εξαίρεση ότι το μηδέν που χρησιμοποιείται για τον καθορισμό της θέσης ενός δεκαδικού σημείου δεν λαμβάνεται ως σημαντικό.

Αυτός ο ορισμός μπορεί να είναι amφίβολος. Π.χ. πόσα σημαντικά ψηφία υπάρχουν στο 100; Για την αποφυγή σύγχυσης, χρησιμοποιείται η επιστημονική σημειολογία. Έτσι, ο αριθμός 1×10^2 έχει ένα σημαντικό ψηφίο, ενώ ο αριθμός 1.0×10^2 έχει δύο.

2. Σημαντικά ψηφία

Η καταγραφή μιας μέτρησης με τον σωστό αριθμό σημαντικών ψηφίων είναι σημαντική γιατί μπορούμε να δούμε πόσο ακριβής είναι η μέτρηση.

Για παράδειγμα, αν ζυγιστεί ένα δείγμα με έναν ζυγό ακρίβειας ± 0.1 mg, αλλά η μέτρηση καταγραφεί ως 1.762 g αντί of 1.7620 g, τότε μπορεί λανθασμένα να θεωρηθεί ότι η ζύγιση έγινε με ζυγό ακρίβειας ± 1 mg.

Ως γενικός κανόνας, οι μαθηματικοί υπολογισμοί που εμπλέκουν πρόσθεση και αφαίρεση γίνονται ως το δεκαδικό που είναι σημαντικό για όλους τους αριθμούς που συμμετέχουν στον υπολογισμό. Έτσι, το άθροισμα των 135.621, 0.33 και 21.2163 είναι 157.17, μιας και το τελευταίο ψηφίο που είναι σημαντικό για όλους τους αριθμούς είναι το εκατοστό:

$$135.621 + 0.33 + 21.2163 = 157.1673 = 157.17$$

Όταν γίνεται πολλαπλασιασμός ή διαίρεση, ο γενικός κανόνας είναι ότι το αποτέλεσμα περιέχει τον ίδιο αριθμό σημαντικών ψηφίων όπως ο αριθμός με τα λιγότερα σημαντικά ψηφία.

$$\frac{22.91 \times 0.152}{16.302} = 0.21361 = 0.214$$

3. Στρογγυλοποίηση

Αν το δεκαδικό που ακολουθεί το τελευταίο σημαντικό ψηφίο είναι μεγαλύτερο του 5, ο αριθμός στρογγυλοποιείται στο επόμενο υψηλότερο δεκαδικό. Αν είναι μικρότερο του 5, ο αριθμός στρογγυλοποιείται στην τιμή του του τελευταίου σημαντικού ψηφίου:

$$9.47 = 9.5$$

$$9.43 = 9.4$$

Αν το τελευταίο δεκαδικό είναι 5, ο αριθμός στρογγυλοποιείται το πλησιέστερο ζυγό ψηφίο:

$$8.65 = 8.6$$

$$8.75 = 8.8$$

$$8.55 = 8.6$$

Αυτό βασίζεται στην στατιστική πρόβλεψη ότι υπάρχει ίση πιθανότητα το τελευταίο σημαντικό ψηφίο πριν το 5 να είναι ζυγό ή μονό. Δηλαδή, σ' ένα μεγάλο δείγμα υπάρχει ίσος αριθμός ζυγών και μονών ψηφίων που ακολουθούν το 5. Για παράδειγμα, αν είναι απαραίτητα 4 σημαντικά ψηφία, τότε ο 45.365 στρογγυλοποιείται σε 45.36, ενώ αντιθέτως ο 45.3651 σε 45.37.

4. Μονάδες έκφρασης συγκέντρωσης

Συγκέντρωση είναι μια γενική μέτρηση που δηλώνει την ποσότητα της διαλυμένης ουσίας σε μια γνωστή ποσότητα διαλύματος. Αν και οι όροι «διαλυμένη ουσία» και «διάλυμα» συχνά σχετίζονται με υγρά δείγματα, οι έννοιες μπορούν να επεκταθούν και σε δείγματα αέριας ή στερεής φάσης. Οι μονάδες έκφρασης της συγκέντρωσης εξαρτώνται από το πως μετρήθηκαν οι ποσότητες της διαλυμένης ουσίας και του διαλύματος.

Molarity (M): Είναι τα mols της διαλυμένης ουσίας ανά λίτρο διαλύματος (mol L^{-1}). Οι μοριακές συγκεντρώσεις χρησιμοποιούνται πολύ συχνά, γι' αυτό και έχει καθιερωθεί συμβολική σημειολογία προς απλοποίηση των εκφράσεων. Η χρήση τετράγωνων αγκύλων υποδεικνύει την μοριακή συγκέντρωση.

Π.χ. 1 M Na^+ μπορεί να συμβολιστεί $[\text{Na}^+] = 1 \text{ mol L}^{-1}$.

Molality (m): Είναι τα mols μιας ουσίας ανά χιλιόγραμμο διαλύτη. Χρησιμοποιείται σε θερμοδυναμικούς υπολογισμούς, όπου υπάρχει ανάγκη μιας μονάδας συγκέντρωσης ανεξάρτητης από την θερμοκρασία. Η molarity βασίζεται σε όγκο διαλύματος, μέσα στον οποίο είναι διαλυμένη η ουσία. Εφόσον ο όγκος εξαρτάται από την θερμοκρασία, η molarity θ' αλλάξει συναρτήσει της θερμοκρασίας. Χρησιμοποιώντας την μάζα του διαλύτη, η έκφραση της συγκέντρωσης είναι ανεξάρτητη της θερμοκρασίας.

4. Μονάδες έκφρασης συγκέντρωσης

Μάζα, όγκος και λόγος μάζα-προς-όγκο

Επί τοις εκατό κατά βάρος (% w/w), επί τοις εκατό κατ' όγκο (% v/v) και επί τοις εκατό μάζα-προς-όγκο (% w/v) εκφράζουν συγκέντρωση μονάδων διαλυμένης ουσίας ανά 100 μονάδες δείγματος. Ένα διάλυμα το οποίο έχει συγκέντρωση 23% w/v περιέχει 23 g διαλυμένης ουσίας ανά 100 mL διαλύματος.

Μέρη ανά εκατομμύριο (parts per million – ppm) και **μέρη ανά δισεκατομμύριο** (parts per billion – ppb) είναι αναλογίας μάζας γραμμαρίων διαλυμένης ουσίας σε ένα εκατομμύριο ή δισεκατομμύριο γραμμάρια δείγματος, αντίστοιχα. Για παράδειγμα, αν ένα διάλυμα Na⁺ είναι 15 ppm, τότε περιέχει 15 mg Na⁺ ανά 1000 g διαλύματος. Επειδή η πυκνότητα ενός τόσο αραιού διαλύματος είναι περίπου 1.00 g/mL, τότε η συγκέντρωση μπορεί να εκφραστεί ως εξής:

$$\text{ppm} = \frac{\text{mg}}{\text{liter}} = \frac{\mu\text{g}}{\text{mL}}$$
$$\text{ppb} = \frac{\mu\text{g}}{\text{liter}} = \frac{\text{ng}}{\text{mL}}$$

Common Units for Reporting Concentration

Name	Units ^a	Symbol
molarity	$\frac{\text{moles solute}}{\text{liters solution}}$	M
molality	$\frac{\text{moles solute}}{\text{kg solvent}}$	<i>m</i>
weight %	$\frac{\text{g solute}}{100 \text{ g solution}}$	% w/w
volume %	$\frac{\text{mL solute}}{100 \text{ mL solution}}$	% v/v
weight-to-volume %	$\frac{\text{g solute}}{100 \text{ mL solution}}$	% w/v
parts per million	$\frac{\text{g solute}}{10^6 \text{ g solution}}$	ppm
parts per billion	$\frac{\text{g solute}}{10^9 \text{ g solution}}$	ppb

$$\text{Concentration} = \frac{\text{amount of solute}}{\text{amount of solution}}$$

5. Είδη σφαλμάτων

Η κάθε μέτρηση έχει μια αβεβαιότητα, η οποία ονομάζεται **πειραματικό σφάλμα**. Τα συμπεράσματα μπορούν να εκφραστούν με υψηλό ή χαμηλό βαθμό εμπιστοσύνης, αλλά ποτέ με απόλυτη βεβαιότητα. Τα πειραματικά σφάλματα κατατάσσονται είτε ως συστηματικά είτε ως τυχαία.

Συστηματικό σφάλμα (systematic error)

Ονομάζεται επίσης και σφάλμα προσδιορισμού. Προέρχεται από κάποιο ελάττωμα στον εξοπλισμό ή τον πειραματικό σχεδιασμό. Εάν ένα πείραμα γίνει με ακριβώς τον ίδιο τρόπο, το σφάλμα είναι αναπαραγώγιμο. Ένα τέτοιο σφάλμα μπορεί να βρεθεί και να διορθωθεί, αλλά δεν είναι εύκολο.

Για παράδειγμα, ένα πεχάμετρο που έχει ρυθμιστεί λανθασμένα παράγει ένα συστηματικό σφάλμα. Αν η ρύθμιση ήταν σε pH 7.00 αντί 7.08 (πραγματική τιμή), τότε κάθε μέτρηση θα είναι 0.08 μονάδες pH χαμηλότερη. Η εξεύρεση του σφάλματος μπορεί να γίνει χρησιμοποιώντας ένα άλλο ρυθμιστικό διάλυμα γνωστού pH.

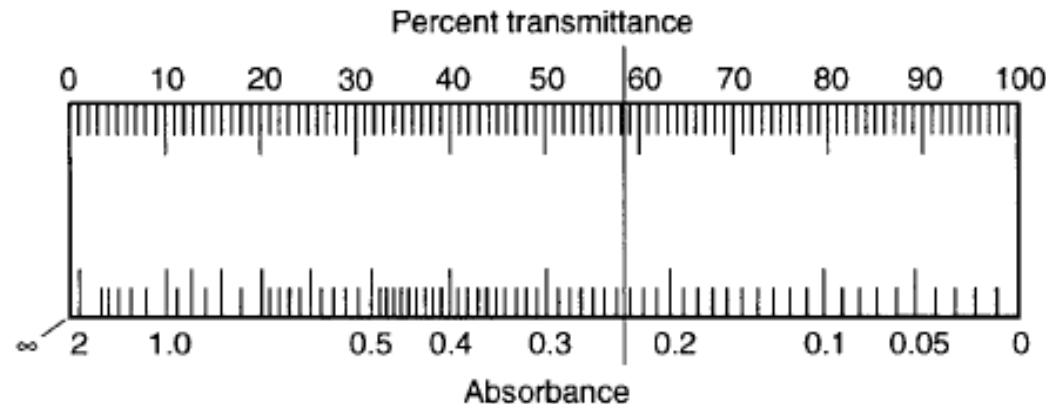
5. Είδη σφαλμάτων

Τυχαίο σφάλμα (random error)

Ονομάζεται και απροσδιόριστο σφάλμα. Προέρχεται από ανεξέλεγκτες μεταβλητές στην μέτρηση. Υπάρχουν ίσες πιθανότητες να είναι θετικό ή αρνητικό.

Υπάρχει τυχαίο σφάλμα στην ανάγνωση μιας κλίμακας. Διαφορετικοί αναλυτές μπορεί να αναγνώσουν διαφορετική μέτρηση στο σχήμα που φαίνεται παρακάτω, η οποία αντιπροσωπεύει διαφορετική παρεμβολή μεταξύ των τιμών.

Ένα άλλο τυχαίο σφάλμα μπορεί να προέρχεται από τον ηλεκτρικό θόρυβο μιας συσκευής. Θετικές και αρνητικές διακυμάνσεις συμβαίνουν με την ίδια συχνότητα και δεν μπορούν ν' αποκλειστούν πλήρως.



6. Ορθότητα και ακρίβεια

Η ορθότητα (precision) περιγράφει την αναπαραγωγιμότητα ενός αποτελέσματος. Εάν μετρηθεί μια ποσότητα αρκετές φορές και υπάρχει μεγάλη εγγύτητα των τιμών, η μέτρηση είναι ορθή. Εάν οι τιμές κυμαίνονται ευρέως, δεν υπάρχει ορθότητα.

Η ακρίβεια (accuracy) περιγράφει πόσο κοντά είναι η μετρούμενη τιμή στην «πραγματική» τιμή. Εάν υπάρχει διαθέσιμο ένα πρότυπο, ακρίβεια είναι η εγγύτητα της μέτρησης με την γνωστή τιμή.

Μια μέτρηση μπορεί να είναι αναπαραγώγιμη, αλλά λανθασμένη. Για παράδειγμα, εάν παρασκευαστεί λάθος ένα διάλυμα για τιτλοδότηση, μπορεί να γίνουν πολλές αναπαραγώγιμες ογκομετρήσεις, αλλά το αποτέλεσμα θα είναι λάθος. Σ' αυτήν την περίπτωση η ορθότητα είναι καλή, αλλά η ακρίβεια φτωχή.

Αντιστρόφως, είναι δυνατόν να παραχθούν αποτελέσματα χαμηλής αναπαραγωγιμότητας γύρω από την σωστή τιμή. Σ' αυτήν την περίπτωση, η ορθότητα είναι φτωχή, αλλά η ακρίβεια καλή.

Μια ιδανική διαδικασία είναι αμφότερα ορθή και ακριβής.

Improving Trueness
Decreasing Systematic Errors

Improving Accuracy
Decreasing Uncertainty



Accurate
Precise



Not Accurate
Precise



Accurate
Not Precise



Not Accurate
Not Precise



Improving Precision
Decreasing Random Errors

7. Απόλυτη και σχετική αβεβαιότητα

Η απόλυτη αβεβαιότητα (absolute uncertainty) εκφράζει το περιθώριο της αβεβαιότητας που σχετίζεται με μια μέτρηση. Εάν η εκτιμώμενη αβεβαιότητα σε μια μέτρηση μιας βαθμονομημένης προχοϊδας είναι ± 0.02 mL, τότε μπορεί να ειπωθεί ότι ± 0.02 mL είναι η απόλυτη αβεβαιότητα που σχετίζεται με την μέτρηση.

Η σχετική αβεβαιότητα (relative uncertainty) συγκρίνει το μέγεθος της απόλυτης αβεβαιότητας με το μέγεθος της σχετιζόμενης μέτρησης. Αν η απόλυτη αβεβαιότητα μιας μέτρησης με προχοϊδα είναι 12.35 ± 0.02 mL, τότε η σχετική αβεβαιότητα θα είναι:

Relative uncertainty:

$$\text{Relative uncertainty} = \frac{\text{absolute uncertainty}}{\text{magnitude of measurement}}$$

$$= \frac{0.02 \text{ mL}}{12.35 \text{ mL}} = 0.002$$

The percent relative uncertainty is simply

Percent relative uncertainty:

$$\text{Percent relative uncertainty} = 100 \times \text{relative uncertainty}$$

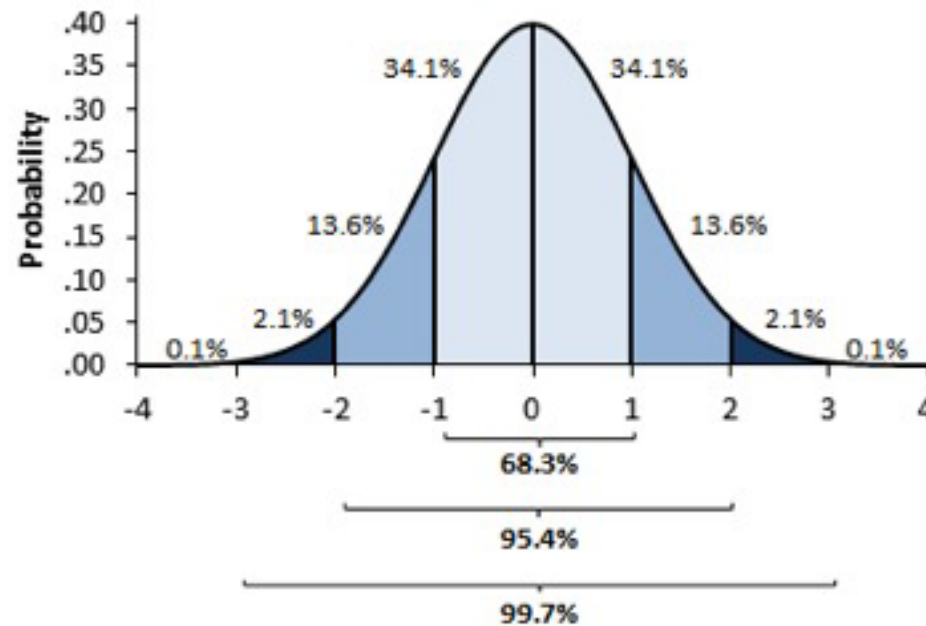
$$= 100 \times 0.002 = 0.2\%$$

Αν η απόλυτη αβεβαιότητα μιας μέτρησης με προχοϊδα είναι σταθερή σε ± 0.02 mL, η % σχετική αβεβαιότητα είναι 0.2% για έναν όγκο 10 mL και 0.1% για 20 mL.

8. Κατανομή κατά Gauss (Gaussian distribution)

Εάν ένα πείραμα επαναληφθεί πολλές φορές και εάν τα σφάλματα είναι τυχαία, τότε τα αποτελέσματα θα τείνουν να συσσωρευθούν γύρω από μια μέση τιμή. Όσα περισσότερα πειράματα πραγματοποιούνται, τόσο περισσότερα τα αποτελέσματα προσεγγίζουν μια ιδανική κατανομή σχήματος κώδωνα (καμπάνας), η οποία ονομάζεται **κατανομή κατά Gauss**.

Γενικά, δεν είναι δυνατόν να πραγματοποιηθούν τόσα πολλά πειράματα σε εργαστηριακή κλίμακα. Συνήθως πραγματοποιούνται 3 έως 5 μετρήσεις, από τις οποίες γίνεται εκτίμηση για ένα ευρύτερο σύνολο.



9. Μέση τιμή και σταθερά απόκλιση

Η καμπύλη Gauss χαρακτηρίζεται από δύο παραμέτρους:

1. **τον αριθμητικό μέσο** $\bar{x}$ (ονομάζεται και μέσος όρος) (mean value or average). Είναι το άθροισμα των μετρούμενων τιμών δια τον αριθμό των μετρήσεων n .

$$\bar{x} = \frac{\sum_i x_i}{n}$$

Όπου x_i είναι η κάθε ξεχωριστή μετρούμενη τιμή. Όπου $\sum_i x_i$ είναι το άθροισμα όλων των μετρούμενων τιμών $x_1 + x_2 + x_3 + \dots + x_n$.

9. Μέση τιμή και σταθερά απόκλιση

2. **την σταθερά απόκλιση s** (standard deviation). Μετράει την εγγύτητα κατανομής των μετρήσεων γύρω από τον μέσο όρο. Όσο πιο μικρή η σταθερά απόκλιση, τόσο οι μετρήσεις συσσωρεύονται εγγύτερα του μέσου όρου.

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n - 1}}$$

Για ένα πολύ μεγάλο σύνολο μετρήσεων, ο μέσος όρος και η σταθερά απόκλιση συμβολίζονται ως μ και σ , αντίστοιχα. Δεν είναι δυνατόν να μετρήσουμε τις παραμέτρους μ και σ , αλλά οι τιμές $\bar{x}$ και s προσεγγίζουν τις πραγματικές μ και σ καθώς ο αριθμός των πειραμάτων αυξάνει.

- Η ποσότητα $n - 1$ στην παραπάνω εξίσωση ονομάζεται **βαθμοί ελευθερίας** (degrees of freedom).
- Το τετράγωνο της σταθεράς απόκλισης ονομάζεται **διακύμανση** (variance).
- Η σταθερά απόκλιση εκπεφρασμένη ως % της μέσης τιμής ($100 \times s/\bar{x}$) ονομάζεται **σχετική σταθερά απόκλιση** (relative standard deviation) ή **συντελεστής διακύμανσης** (coefficient of variation).

Βιβλιογραφία

- Christian G.D., Dasgupta P.K., Schug K.A., **2014**. Analytical Chemistry, 7th Edition, Wiley, NJ, U.S.A.
- Harris D.C., **2016**. Quantitative Chemical Analysis, 9th Edition, W. H. Freeman and Company, NY, U.S.A.