

Κεφάλαιο 4: Στατιστική

Ενόργανη Χημική Ανάλυση (ΜΠ 212)

Καθ. Έλια Ψυλλάκη

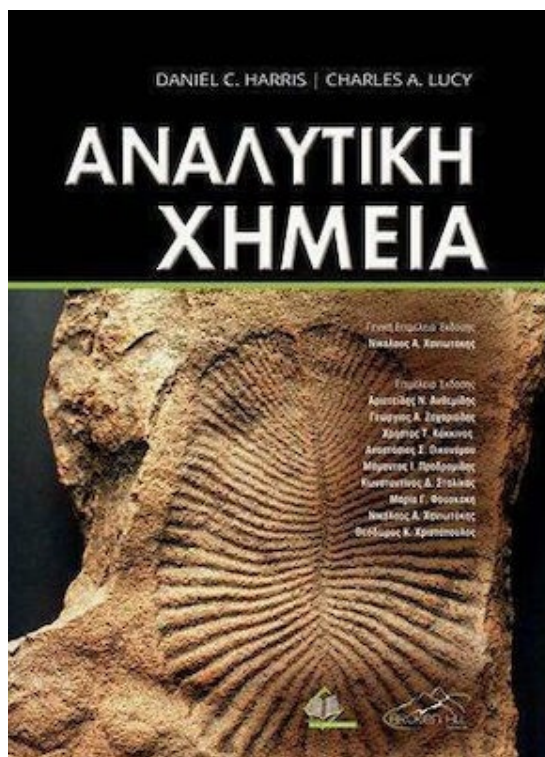
Εργαστήριο Υδατικής Χημείας
Σχολή Χημικών Μηχανικών και
Μηχανικών Περιβάλλοντος
Πολυτεχνείο Κρήτης

Γραφείο διδάσκουσας: Κ1.102

Τηλ: 2821037810

email: epsillakis@tuc.gr

Κωδικοί	Υποχρεωτικά Μαθήματα	(Θ-A-E)	Ω	ECTS	Διδάσκων
ΜΠ 212	Ενόργανη Χημική Ανάλυση	(3-1-2)	6	4	Ε. Ψυλλάκη



Τίτλος: **Αναλυτική Χημεία**

Συγγραφείς: **Daniel C. Harris, Charles A. Lucy**

Στοιχεία έκδοσης: Χρόνος έκδ. 2021, σελ. 900, ISBN: 9789925576111

Έκδοση-Διάθεση: Broken Hill Publishers Ltd

(Quantitative Chemical Analysis, Daniel C. Harris and Charles A. Lucy, © 2020 W. H. Freeman and Company)

Στατιστική

Κεφάλαιο 4

Περίγραμμα κεφαλαίου

- 4-1 Κατανομή Gauss
- 4-2 Σύγκριση τυπικών αποκλίσεων με τη δοκιμασία F (F Test)
- 4-3 Διαστήματα εμπιστοσύνης
- 4-4 Σύγκριση μέσων τιμών με τη δοκιμή Student's t
- 4-5 *Δοκιμές t με υπολογιστικό φύλλο*
- 4-6 Δοκιμή Grubbs για αποκλίνουσες τιμές
- 4-7 Η μέθοδος των ελαχίστων τετραγώνων
- 4-8 Καμπύλες βαθμονόμησης
- 4-9 Υπολογιστικό φύλλο για τη μέθοδο ελαχίστων τετραγώνων

ΕΙΝΑΙ ΣΩΣΤΗ Η ΜΕΤΡΗΣΗ ΤΗΣ ΓΛΥΚΟΖΗΣ ΣΤΟ ΑΙΜΑ ΜΟΥ;

Μεγάλες μεταβολές μπορεί να σημαίνουν

- Προβλήματα υγείας
- Κακή χρήση
- Μετρητή που δυσλειτουργεί



Οι κατασκευαστές μετρητών γλυκόζης συνιστούν την ανάλυση ενός διαλύματος γνωστής συγκέντρωσης γλυκόζης για έλεγχο

Μετρήσεις δείγματος ελέγχου	Συγκέντρωση δείγματος ελέγχου
90 mg/dL	99 mg/dL
94 mg/dL	
85 mg/dL	
95 mg/dL	
93 mg/dL	
Μέση τιμή = 91.4 mg/dL	

Όλες οι μετρήσεις εμπεριέχουν πειραματική αβεβαιότητα και η στατιστική μάς παρέχει εργαλεία για να:

- Αποδεχτούμε συμπεράσματα που έχουν μεγάλη πιθανότητα να είναι σωστά
- Να απορρίψουμε συμπεράσματα που έχουν μικρή πιθανότητα να είναι σωστά

(Παρακάτω θα δούμε: 1.3% πιθανότητα η μετρούμενη να απέχει από την πραγματική)

Μέση τιμή και τυπική απόκλιση

Η **αριθμητική μέση τιμή**, $\bar{x}$, είναι το άθροισμα όλων των τιμών που μετρήθηκαν διαιρεμένων με το n , τον αριθμό των μετρήσεων:

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

καθώς το n αυξάνεται, $\bar{x} \rightarrow \mu (\infty)$

μ =πραγματική τιμή

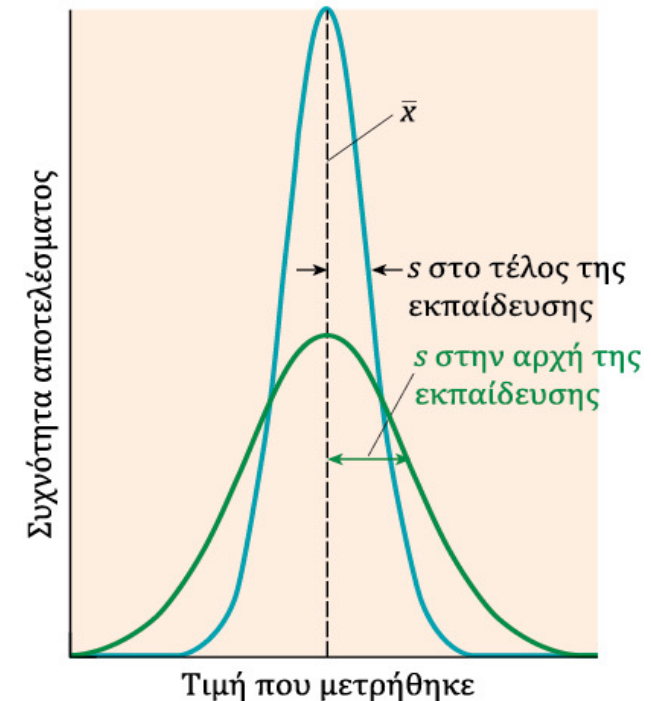
Η **τυπική απόκλιση**, s , μετράει το πόσο κοντά στη μέση τιμή συγκεντρώνονται τα δεδομένα

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

καθώς το n αυξάνεται, $s \rightarrow \sigma (\infty)$

σ =τυπική απόκλιση πληθυσμού

$n-1$ =βαθμοί ελευθερίας



EIKONA 4-2 Υποθετικές καμπύλες Gauss για δύο σειρές εργαστηριακών αναλύσεων φοιτητών, όπου η μία έχει διπλάσια τυπική απόκλιση από την άλλη. Ο αριθμός των μετρήσεων που περιλαμβάνει κάθε καμπύλη περιγράφεται από το εμβαδόν κάτω από την κάθε καμπύλη, και σε αυτή την περίπτωση είναι ίδιος.

Τυπική απόκλιση, s

Όσο μικρότερη είναι η τυπική απόκλιση, τόσο περισσότερο τα δεδομένα είναι συγκεντρωμένα γύρω από τη μέση.

Πιστότητα: επαναληψιμότητα / αναπαραγωγικότητα

Ορθότητα: εγγύτητα προς την «πραγματική» τιμή (μ)

- Ένα πείραμα που παρουσιάζει μικρότερη τυπική απόκλιση είναι πιο επαναλήψιμο από ένα που παρουσιάζει μεγαλύτερη τυπική απόκλιση.
- Υψηλότερη πιστότητα δεν συνεπάγεται αναγκαστικά υψηλότερη ορθότητα.



κακή ορθότητα,
άριστη ακρίβεια

άριστη ορθότητα,
κακή ακρίβεια

άριστη ορθότητα,
άριστη ακρίβεια

ΜΠ212-Ενοργάνη Χημική Ανάλυση-Κεφάλαιο 4

2.13 (3.5)

measurement accuracy

accuracy of measurement
accuracy

closeness of agreement between a **measured quantity value** and a **true quantity value** of a measurand

2.15

measurement precision

precision

closeness of agreement between **indications** or **measured quantity values** obtained by replicate **measurements** on the same or similar objects under specified conditions

Άλλες στατιστικές παράμετροι

- **Βαθμοί ελευθερίας** = $n - 1$ (degrees of freedom)
- Το τετράγωνο της τυπικής απόκλισης (σ^2) ονομάζεται **διακύμανση** (variance).
- Τυπική απόκλιση εκφραζόμενη ως % ποσοστό της μέσης τιμής ($= 100 s / \bar{x}$) ονομάζεται **σχετική τυπική απόκλιση ή συντελεστής διακύμανσης** (RSD; relative standard deviation, coefficient of variation)

Παράδειγμα:

Να υπολογίσετε τη μέση τιμή, την τυπική απόκλιση και τη σχετική τυπική απόκλιση των τιμών 821, 783, 834 και 855

Λύση: Η μέση τιμή είναι

$$\bar{x} = \frac{821 + 783 + 834 + 855}{4} = 823.2$$

$$\bar{x} \pm s = 823.2 \pm 30.3 \text{ (n=4)}$$

Για να αποφύγουμε τη συσσώρευση σφαλμάτων στρογγυλοποίησης, διατηρούμε στη μέση τιμή ένα ψηφίο περισσότερο από αυτό που υπάρχει στα πρωταρχικά δεδομένα.

Η τυπική απόκλιση είναι

$$s = \sqrt{\frac{(821 - 823.2)^2 + (783 - 823.2)^2 + (834 - 823.2)^2 + (855 - 823.2)^2}{(4 - 1)}} = 30.3$$

$$\text{Relative standard deviation} = 100 \times \frac{s}{\bar{x}} = 100 \times 30.3/823.2 = 3.7\%$$

Τι θα γίνει εάν διαιρεθούν όλες οι τιμές δια 2?

Υπολογιστικά φύλλα για μέση και τυπική απόκλιση

Τα υπολογιστικά φύλλα έχουν ενσωματωμένες συναρτήσεις

- Μέση τιμή:

=AVERAGE(B1:B4)

- Τυπική απόκλιση:

=STDEV.S(B1:B4)

	A	B
1	Excel	821
2		783
3		834
4		855
5	Average =	823.25
6	Std dev =	30.27
7	B5 = AVERAGE(B1:B4)	
8	B6 = STDEV(B1:B4)	

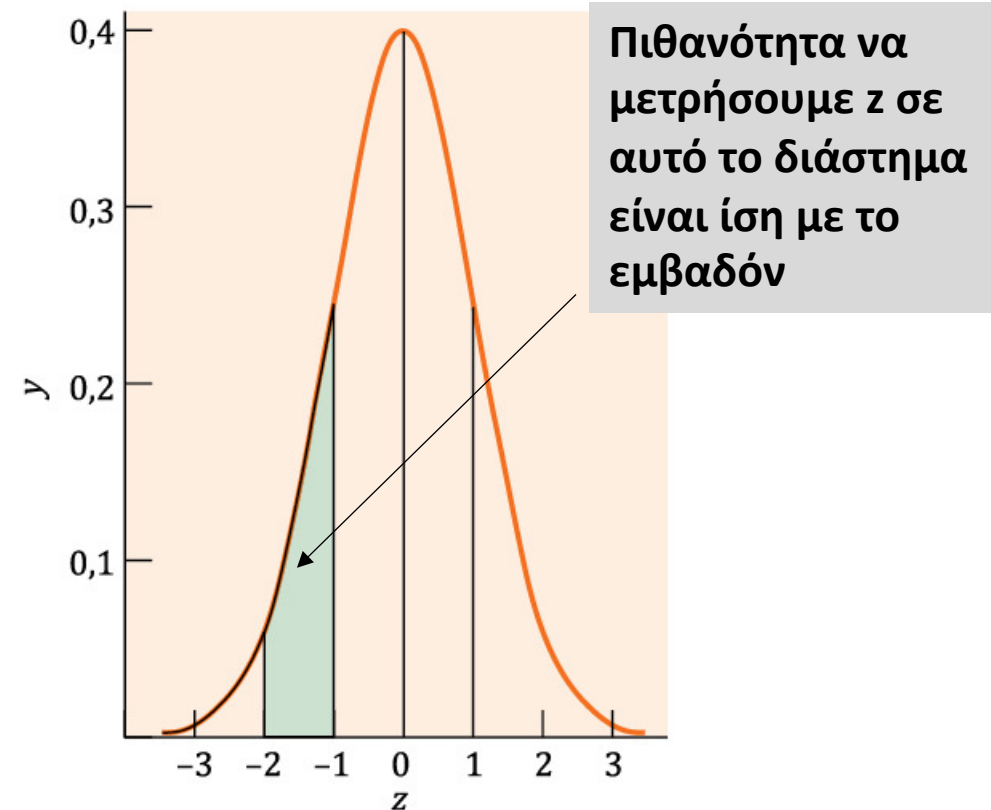
Τυπική απόκλιση και πιθανότητα

Συνάρτηση καμπύλης Gauss

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

Για πεπερασμένο αριθμό μετρήσεων $\mu \approx \bar{x}$ και $\sigma \approx s$.
Είναι χρήσιμο να εκφράζουμε αποκλίσεις από τη μέση τιμή σε **πολλαπλάσια, z, της τυπικής απόκλισης**.
Δηλαδή μετατρέπουμε το x σε z, με τον τύπο

$$z \approx \frac{x - \mu}{\sigma} \approx \frac{x - \bar{x}}{s}$$



EIKONA 4-3 Καμπύλη Gauss όπου $\mu = 0$ και $\sigma = 1$. Καμπύλη Gauss της οποίας το εμβαδόν είναι 1 ονομάζεται κανονική καμπύλη. Η τετμημένη $z = (x - \mu)/\sigma$ είναι η απόσταση από τη μέση τιμή μετρημένη σε μονάδες τυπικής απόκλισης. Όταν $z = 2$, βρισκόμαστε σε απόσταση 2 τυπικών αποκλίσεων από τη μέση τιμή.

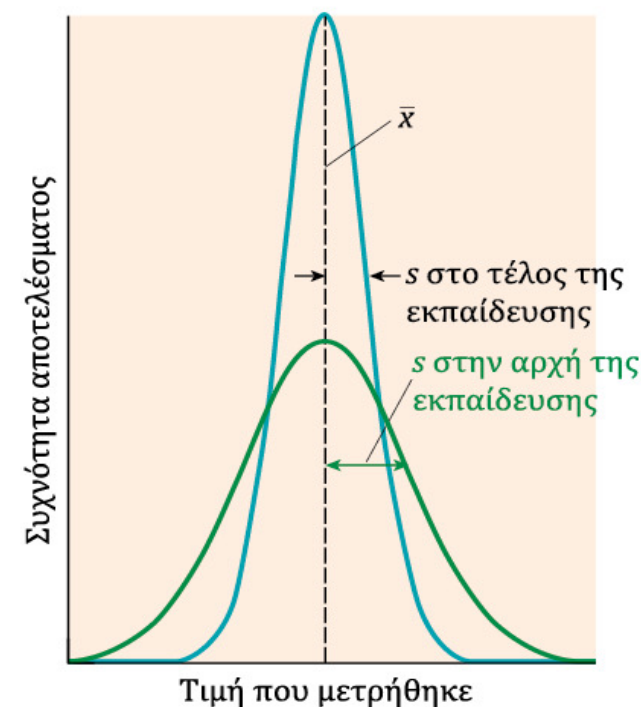
Εφόσον το άθροισμα των πιθανοτήτων όλων των μετρήσεων πρέπει να είναι 1.

- Το εμβαδόν της περιοχής κάτω από όλη την καμπύλη από το $z = -\infty$ έως το $z = +\infty$ πρέπει να είναι ίσο με τη μονάδα.

Η τυπική απόκλιση είναι μέτρο του εύρους της καμπύλης Gauss.

- Όσο μεγαλύτερη είναι η τιμή του σ , τόσο πλατύτερη είναι η καμπύλη.

Εύρος	Ποσοστό μετρήσεων
$\mu \pm 1\sigma$	68.3
$\mu \pm 2\sigma$	95.5
$\mu \pm 3\sigma$	99.7



ΕΙΚΟΝΑ 4-2 Υποθετικές καμπύλες Gauss για δύο σειρές εργαστηριακών αναλύσεων φοιτητών, όπου η μία έχει διπλάσια τυπική απόκλιση από την άλλη. Ο αριθμός των μετρήσεων που περιλαμβάνει κάθε καμπύλη περιγράφεται από το εμβαδόν κάτω από την κάθε καμπύλη, και σε αυτή την περίπτωση είναι ίδιος.

Τυπική απόκλιση της μέσης τιμής

- Αβεβαιότητα περιγράφεται από S_x
- Όσες περισσότερες φορές μετράμε ένα μέγεθος, τόσο περισσότερη εμπιστοσύνη έχουμε η μέση τιμή μας να είναι κοντά στη μέση τιμή του πληθυσμού.
- **Τυπική απόκλιση της μέσης τιμής** (ή τυπική αβεβαιότητα ή τυπικό σφάλμα) για σειρές n μετρήσεων, u_x :

$$u_x = \frac{S_x}{\sqrt{n}}$$

- Το S_x πλησιάζει μια σταθερή τιμή καθώς το n πλησιάζει το ∞ .
- Το u_x πλησιάζει το 0 καθώς το n πλησιάζει το ∞ .
- Η αβεβαιότητα μειώνεται αναλογικά με το $\sqrt{n}$

4-2 Σύγκριση τυπικών αποκλίσεων με τη δοκιμασία F (F Test)

«Σε δύο σειρές μετρήσεων, είναι οι μέσες τιμές «στατιστικά διαφορετικές» συνυπολογίζοντας την πειραματική αβεβαιότητα;».

- Αποδοχή συμπερασμάτων με μεγάλη πιθανότητα να είναι σωστά.
- Απόρριψη συμπερασμάτων με μεγάλη πιθανότητα να είναι εσφαλμένα

Η μηδενική υπόθεση χρησιμοποιείται με διάφορες δοκιμές. Υπόθεση ότι δύο σειρές δεδομένων προέρχονται από πληθυσμούς με ίδιες ιδιότητες δηλαδή **τυπική απόκλιση σ , (δοκιμασία F) ή μέση τιμή μ (δοκιμασία t)**. Τυχόν διαφορές προέρχονται μόνο από τη μεταβλητότητα στις μετρήσεις. Απορρίπτουμε τη μηδενική υπόθεση αν υπάρχει πιθανότητα $<5\%$ οι δύο σειρές πειραματικών αποτελεσμάτων να προέρχονται από δύο πληθυσμούς με την ίδια τυπική απόκλιση πληθυσμού.

Η δοκιμασία F συγκρίνει την τυπική απόκλιση

Θα εξετάσουμε τη **μηδενική υπόθεση** που εξετάζει αν δύο σειρές μετρήσεων προέρχονται από πληθυσμούς με ίδια τυπική απόκλιση πληθυσμού (σ).

- Τυχόν παρατηρούμενες διαφορές θα προέρχονται μόνο από την μεταβλητότητα στις μετρήσεις.
- Απορρίπτουμε τη μηδενική υπόθεση αν υπάρχει πιθανότητα $<5\%$ οι δύο σειρές πειραματικών αποτελεσμάτων να προέρχονται από δύο πληθυσμούς με την ίδια τυπική απόκλιση πληθυσμού.

F-Test: Σύγκριση τυπικής απόκλισης

Οι τυπικές αποκλίσεις των δύο σειρών
είναι «στατιστικά διαφορετικές».

$$F_{\text{calculated}} = \frac{s_1^2}{s_2^2} \quad \text{όπου } s_1 \geq s_2$$

*Βάζουμε τη μεγαλύτερη
τυπική απόκλιση στον
αριθμητή ώστε $F \geq 1$.*

If $F_{\text{calculated}} > F_{\text{table}}$

- Υπάρχει <5% πιθανότητα οι δύο σειρές να προέρχονται από πληθυσμούς με την ίδια τυπική απόκλιση πληθυσμού.
- Η διαφορά θεωρείται σημαντική.

ΠΙΝΑΚΑΣ 4-3 Κρίσιμες τιμές $F = s_1^2 / s_2^2$ σε επίπεδο εμπιστοσύνης 95% για δίπλευρη δοκιμασία F

Βαθμοί ελευθερίας του s_2	Βαθμοί ελευθερίας του s_1													
	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
2	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.50
3	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.08	13.90
4	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.46	8.26
5	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.23	6.02
6	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.07	4.85
7	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.36	4.14
8	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.89	3.67
9	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.56	3.33
10	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.31	3.08
11	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.12	2.88
12	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	2.96	2.72
13	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.84	2.60
14	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.73	2.49
15	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.64	2.40
16	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.57	2.32
17	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.50	2.25
18	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.44	2.19
19	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.39	2.13
20	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.35	2.09
30	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.07	1.79
∞	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.57	1.00

Μετρήσεις HCO_3^- σε αίμα αλόγου

Κάποιοι εκπαιδευτές κάνουν ενέσεις με NaHCO_3 στα άλογα πριν από μια ιπποδρομία για να εξουδετερώσουν το γαλακτικό οξύ που συσσωρεύεται κατά τη διάρκεια της έντονης μυϊκής δραστηριότητας. Για να επιβληθεί η απαγόρευση αυτής της πρακτικής, μετράται η συγκέντρωση HCO_3^- στο αίμα των αλόγων μετά από την ιπποδρομία.

Οι αρχές έπρεπε να πιστοποιήσουν ένα νέο όργανο:

	Αρχικό όργανο	Νέο όργανο
Μέση τιμή ($\bar{x}$, mM)	36.14	36.20
Τυπική απόκλιση (s, mM)	0.28	0.47
Αριθμός μετρήσεων (n)	10	4

Λύση: Υπολογίζουμε $F_{\text{calculated}}$ με την Εξίσωση 4-6:

$$F_{\text{calc}} = \frac{s_1^2}{s_2^2} = \frac{(0.47)^2}{(0.28)^2} = 2.82$$

Στον Πίνακα 4-3 βρίσκουμε $F_{\text{table}} = 5,08$ στη στήλη με 3 βαθμούς για το s_1 (βαθμοί ελευθερίας = $n - 1$) και στην σειρά για 9 βαθμούς ελευθερίας για το s_2 .

Εφόσον $F_{\text{calculated}} (= 2,82) < F_{\text{table}} (= 5,08)$, αποδεχόμαστε την μηδενική υπόθεση και συμπεραίνουμε ότι τα s_1 και s_2 δεν είναι σημαντικά διαφορετικά.

Αν υπήρχαν $n = 21$ επαναλήψεις σε κάθε σειρά δεδομένων, θα ήταν σημαντική η διαφορά στις τυπικές αποκλίσεις;
 $F_{\text{calculated}} (= 2,82) > F_{\text{table}} (= 2,46)$

4-3 Διαστήματα εμπιστοσύνης

Υπολογισμός των Διαστημάτων Εμπιστοσύνης

Η δοκιμή **t** του **Student (Student's t test)** είναι ένα στατιστικό εργαλείο που χρησιμοποιείται πολύ συχνά ως μετρό των διαστημάτων εμπιστοσύνης και για να συγκρίνει αποτελέσματα από διαφορετικά πειράματα (**μέση τιμή μ**). Είναι η δοκιμή που θα μπορούσαμε να χρησιμοποιήσουμε για να υπολογίσουμε την πιθανότητα ότι η μέτρηση γλυκόζης του συγγενούς μας συμφωνεί με την «γνωστή» τιμή. Το διάστημα εμπιστοσύνης υπολογίζεται από την εξίσωση

$$\text{Διάστημα εμπιστοσύνης} = \bar{x} \pm \frac{ts}{\sqrt{n}} = \bar{x} \pm t u_x$$

Φυσική σημασία: Αν επαναλαμβάναμε n μετρήσεις πολλές φορές για να υπολογίσουμε τη μέση τιμή και την τυπική απόκλιση, το διάστημα εμπιστοσύνης 95% θα περιλάμβανε την πραγματική μέση τιμή του πληθυσμού (της οποίας δεν γνωρίζουμε την τιμή) στο 95% των σειρών n μετρήσεων.

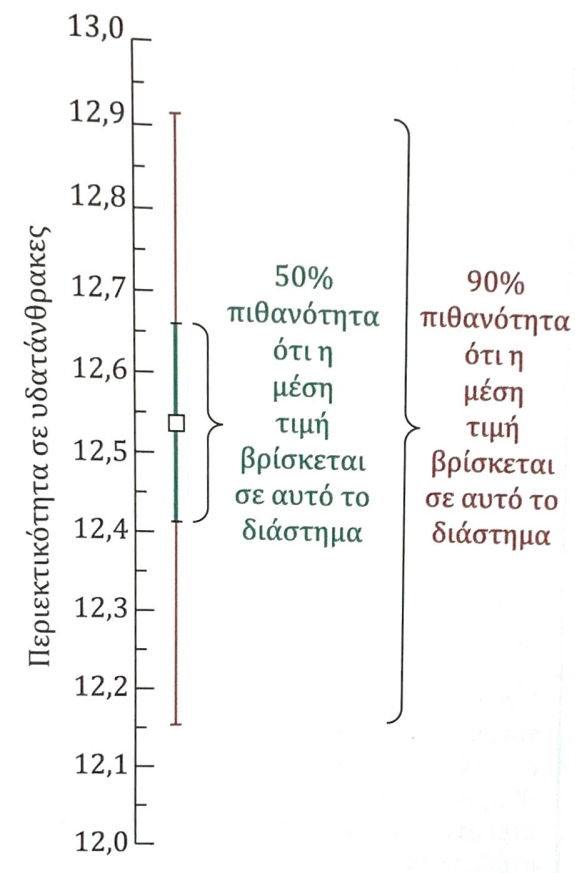
«είμαστε 95% σίγουροι ότι η *πραγματική μέση τιμή* βρίσκεται μέσα στο διάστημα εμπιστοσύνης».

ΠΙΝΑΚΑΣ 4-4 Τιμές Student's t

Βαθμοί ελευθερίας	Στάθμη εμπιστοσύνης (%)						
	50	90	95	98	99	99.5	99.9
1	1.000	6.314	12.706	31.821	63.656	127.321	636.578
2	0.816	2.920	4.303	6.965	9.925	14.089	31.598
3	0.765	2.353	3.182	4.541	5.841	7.453	12.924
4	0.741	2.132	2.776	3.747	4.604	5.598	8.610
5	0.727	2.015	2.571	3.365	4.032	4.773	6.869
6	0.718	1.943	2.447	3.143	3.707	4.317	5.959
7	0.711	1.895	2.365	2.998	3.500	4.029	5.408
8	0.706	1.860	2.306	2.896	3.355	3.832	5.041
9	0.703	1.833	2.262	2.821	3.250	3.690	4.781
10	0.700	1.812	2.228	2.764	3.169	3.581	4.587
15	0.691	1.753	2.131	2.602	2.947	3.252	4.073
20	0.687	1.725	2.086	2.528	2.845	3.153	3.850
25	0.684	1.708	2.060	2.485	2.787	3.078	3.725
30	0.683	1.697	2.042	2.457	2.750	3.030	3.646
40	0.681	1.684	2.021	2.423	2.704	2.971	3.551
60	0.679	1.671	2.000	2.390	2.660	2.915	3.460
120	0.677	1.658	1.980	2.358	2.617	2.860	3.373
∞	0.674	1.645	1.960	2.326	2.576	2.807	3.291

ΠΑΡΑΔΕΙΓΜΑ Υπολογισμός των διαστημάτων εμπιστοσύνης

Η περιεκτικότητα μιας γλυκοπρωτεΐνης (μιας πρωτεΐνης ενωμένης με σάκχαρα) σε υδατάνθρακες βρίσκεται με διαδοχικές αναλύσεις ότι είναι 12.6, 11.9, 13.0, 12.7 και 12.5 %w/w (g υδατανθράκων/100 g γλυκοπρωτεΐνης). Να βρείτε τα διαστήματα εμπιστοσύνης 50% και 90% για την περιεκτικότητα σε υδατάνθρακες.



ΠΑΡΑΔΕΙΓΜΑ Υπολογισμός των διαστημάτων εμπιστοσύνης

Λύση: Πρώτα υπολογίζουμε τα $\bar{x}$ ($= 12,5_4$) και s ($= 0,4_0$) για τις πέντε μετρήσεις.

Για το διάστημα εμπιστοσύνης 50%, βρίσκουμε το t στον Πίνακα 4-4 στη στήλη 50 και στη γραμμή για $n=4$ βαθμούς ελευθερίας. Η τιμή του t είναι 0,741, έτσι το διάστημα εμπιστοσύνης 50% είναι:

$$\text{Διάστημα εμπιστοσύνης 50\%} = \bar{x} + \frac{ts}{\sqrt{n}} = 12.5_4 \pm \frac{(0.741)(0.4_0)}{\sqrt{5}} = 12.5_4 \pm 0.1_3 \text{ wt\%}$$

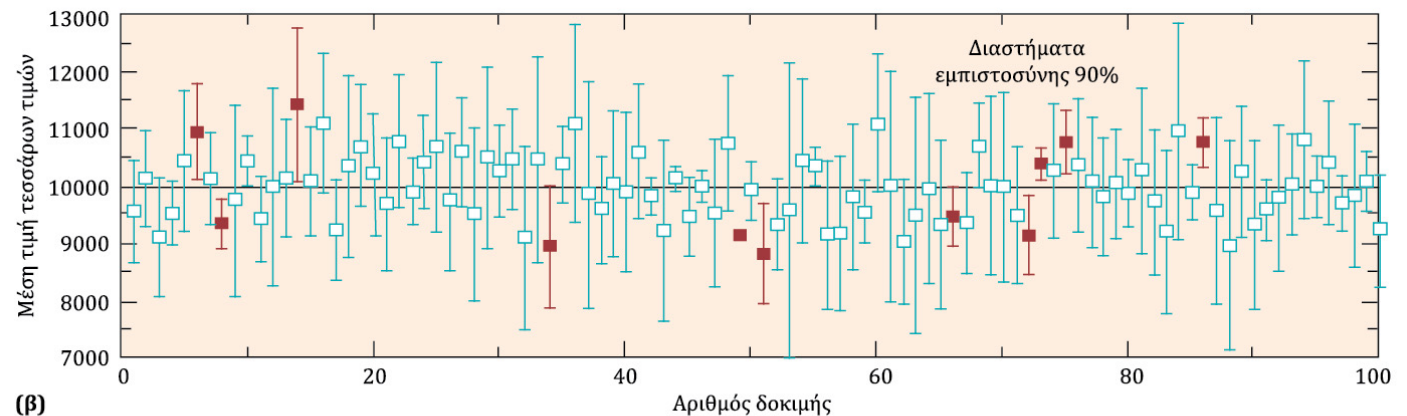
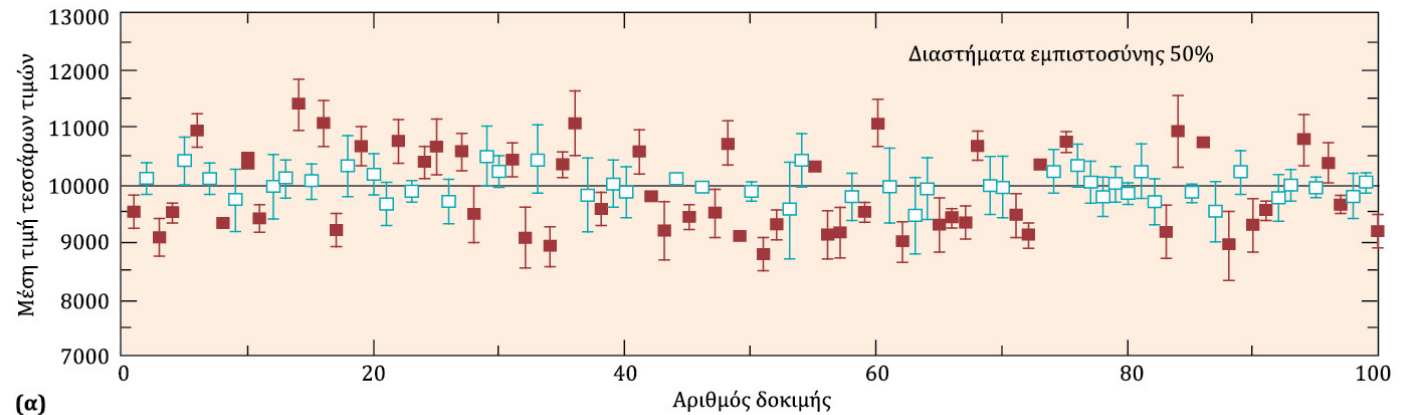
Το διάστημα εμπιστοσύνης 90% είναι

$$\text{Διάστημα εμπιστοσύνης 90\%} = \bar{x} + \frac{ts}{\sqrt{n}} = 12.5_4 \pm \frac{(2.132)(0.4_0)}{\sqrt{5}} = 12.5_4 \pm 0.3_8 \text{ wt\%}$$

Διαστήματα εμπιστοσύνης 50% και 90% για την ίδια σειρά τυχαίων δεδομένων

- Κάθε σημείο αντιπροσωπεύει έναν μέσο όρο $\bar{x}$ ($n = 4$)
- Τιμή πληθυσμού (μ) 10.000

Τα γεμάτα τετράγωνα αντιπροσωπεύουν τα σημεία δεδομένων των οποίων το διάστημα εμπιστοσύνης δεν περιλαμβάνει την πραγματική μέση τιμή του πληθυσμού με τιμή 10.000



EIKONA 4-5 Διαστήματα εμπιστοσύνης 50% και 90% για την ίδια σειρά τυχαίων δεδομένων. Τα γεμάτα τετράγωνα αντιπροσωπεύουν τα σημεία δεδομένων των οποίων το διάστημα εμπιστοσύνης δεν περιλαμβάνει την πραγματική μέση τιμή του πληθυσμού με τιμή 10.000.

Πάντα να δηλώνετε τι είδους αβεβαιότητα αναφέρετε

Μετράμε τον όγκο ενός δοχείου 5 φορές: 6.375, 6.372, 6.364, 6.377 και 6.375 ($\bar{x}=6.3746$, $s=0.0018$)

Μπορείτε να αναφέρετε την αβεβαιότητα ως

Τυπική απόκλιση $\bar{x} \pm s = 6,3746 \pm 0.0018$

ή

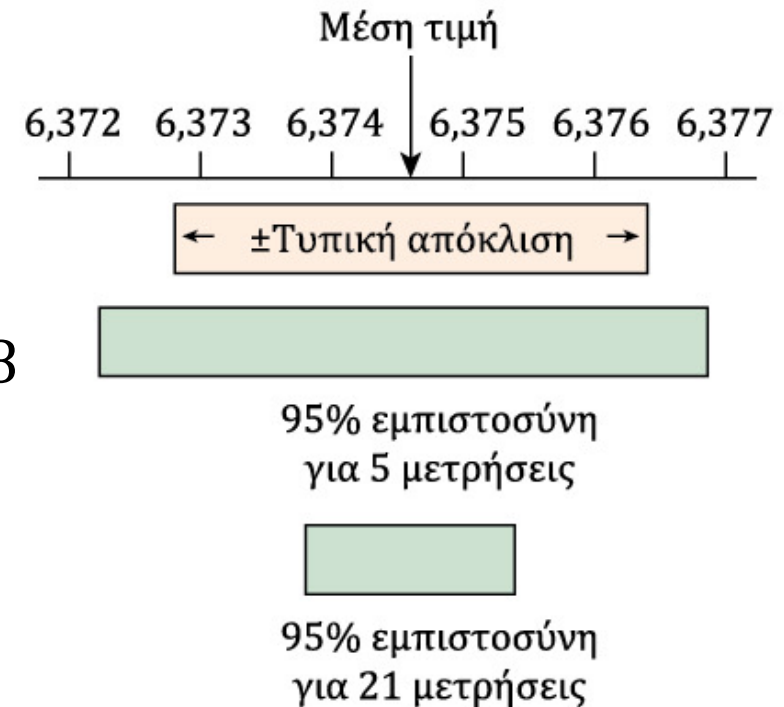
Τυπική απόκλιση μέσης τιμής $\bar{x} \pm s/\sqrt{n} = 6,3746 \pm 0.0008$

ή

Διάστημα εμπιστοσύνης π.χ. 95%

$$\bar{x} \pm \frac{ts}{\sqrt{n}} = 6,3746 \pm 2.776 \times \frac{0.0018}{\sqrt{5}} = 6,3746 \pm 0.0023$$

Μειώνουμε την αβεβαιότητα κάνοντας περισσότερες μετρήσεις.



Υπολογιστικό φύλλο για τον υπολογισμό του διαστήματος εμπιστοσύνης

Βρίσκοντας τα διαστήματα εμπιστοσύνης με το Excel

- Βρίσκουμε τον αριθμό σημείων, n , στο κελί C5 με την συνάρτηση «=COUNT(A4:A10)»:

=COUNT(A4:A10)

- Το όρισμα στο κελί C8 είναι «=T.INV.2T(1-C7,C6)», που μας δίνει το Student's t για στάθμη εμπιστοσύνης 95% και 4 βαθμούς ελευθερίας.

	A	B	C	D	E	F
1	Confidence interval					
2						
3	Data	mean =	6.3746		=AVERAGE(A4:A10)	
4	6.375	stdev =	0.0018		=STDEV.S(A4:A10)	
5	6.372	n =	5		=COUNT(A4:A10)	
6	6.374	degrees of freedom =	4		=C5-1	
7	6.377	confidence level =	0.95		⇐ Enter	
8	6.375	Student's t =	2.776		=T.INV.2T(1-C7,C6)	
9		confidence interval =	0.0023		=C8*C4/SQRT(C5)	
10						

Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

4-4 Σύγκριση μέσων τιμών με τη δοκιμή Student's t

Σύγκριση μέσων τιμών με τη δοκιμή Student's t

Αν πραγματοποιήσουμε δύο σειρές μετρήσεων της ίδιας ποσότητας, οι μέσες τιμές των δύο σειρών γενικά θα διαφέρουν μεταξύ τους λόγω τυχαίων μεταβολών στις μετρήσεις.

Χρησιμοποιούμε τη δοκιμή t (t test) για να αποφασίσουμε αν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των δύο μέσων τιμών $\bar{x}_1$ και $\bar{x}_2$.

- Απορρίπτουμε τη μηδενική υπόθεση αν υπάρχει πιθανότητα μικρότερη $p = 5\%$ οι δύο σειρές δεδομένων να προέρχονται από πληθυσμούς με την ίδια μέση τιμή πληθυσμού.
- Η διαφορά θεωρείται σημαντική.

Σύγκριση μέσων τιμών με τη δοκιμή Student's t

Περίπτωση 1

Σύγκριση $\bar{x}$ με αποδεκτή τιμή:

- Μετράμε μια ποσότητα πολλές φορές.
- Υπολογίζουμε τη μέση τιμή και την τυπική απόκλιση.

Χρειάζεται να συγκρίνουμε την απάντησή μας με μια αποδεκτή τιμή;

Περίπτωση 2

Μετράμε μια ποσότητα πολλές φορές με δύο διαφορετικές μεθόδους.

Τα δύο αποτελέσματα συμφωνούν μεταξύ τους «μέσα στα όρια της πειραματικής αβεβαιότητας»;

Περίπτωση 3 (paired test)

- Το δείγμα A μετράται μια φορά με τη μέθοδο 1 και μια φορά με τη μέθοδο 2 και οι δύο μετρήσεις δεν δίνουν ακριβώς το ίδιο αποτέλεσμα.
- Μετά, ένα διαφορετικό δείγμα B μετράται μια φορά με τη μέθοδο 1 και μια φορά με τη μέθοδο 2.

Οι δύο μέθοδοι συμφωνούν «μέσα στα όρια της πειραματικής αβεβαιότητας»;

Περίπτωση 1. Σύγκριση μετρούμενης τιμής με «αποδεκτή» τιμή

Αγοράσαμε ένα πρότυπο υλικό αναφοράς άνθρακα πιστοποιημένο από το National Institute of Standards and Technology ότι περιέχει θείο 3,19 %w/w. Δοκιμάζουμε μια νέα αναλυτική μέθοδο για να δούμε αν τα αποτελέσματα συμφωνούν με τη γνωστή τιμή. Οι μετρήσεις είναι 3,29, 3,22, 3,30 και 3,23 wt % θείο, με μέση τιμή $\bar{x} = 3,260$ και τυπική απόκλιση $s = 0,041$. **Συμφωνεί η απάντησή μας με τη γνωστή τιμή;**

$$\text{Διάστημα εμπιστοσύνης 95\%} = \bar{x} \pm \frac{ts}{\sqrt{n}} = 3.260 \pm \frac{(3.182)(0.041)}{\sqrt{4}} = 3.260 \pm 0.065$$

- Διάστημα εμπιστοσύνης 95% = 3.19₅ ως 3.32₅ wt%
- Η γνωστή απάντηση (3,19 %w/w) είναι οριακά εκτός του διαστήματος εμπιστοσύνης 95%.
- Συμπεραίνουμε ότι η μέθοδός μας δίνει «διαφορετικό» αποτέλεσμα από τη γνωστή τιμή.

Θα ήταν συνετό να κάνουμε περισσότερες μετρήσεις πριν συμπεράνουμε ότι η μέθοδός μας δεν είναι ακριβής.

n-1	95%
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306
9	2.262
10	2.228
15	2.131
20	2.086
25	2.060
30	2.042
40	2.021
60	2.000
120	1.980
∞	1.960

Περίπτωση 2α: Σύγκριση Πολλαπλών Μετρήσεων

Μετρήσεις HCO_3^- σε αίμα αλόγου

	Αρχικό όργανο	Νέο όργανο
Μέση τιμή ($\bar{x}$, mM)	36.14	36.20
Τυπική απόκλιση (s, mM)	0.28	0.47
Αριθμός μετρήσεων (n)	10	4

Περίπτωση 2α: Σύγκριση Πολλαπλών Μετρήσεων

Οι τυπικές αποκλίσεις είναι $s_1 = 0,28$ και $s_2 = 0,47$ mM, οι οποίες δεν διαφέρουν σημαντικά μεταξύ τους. Η δοκιμή t όταν οι s δεν είναι σημαντικά διαφορετικές:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{\text{pooled}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{|36.14 - 35.20|}{0.33_8} \sqrt{\frac{10 \times 4}{10 + 4}} = 0.30_0$$

$$s_{\text{pooled}} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}} = \sqrt{\frac{0.28_1(10 - 1) + 0.47_2(4 - 1)}{10 + 4 - 2}} = 0.33_8$$

$(n_1 + n_2 - 2) = 12$
βαθμούς ελευθερίας
βρίσκεται μεταξύ του
2,228 και του 2,131

n-1	95%
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306
9	2.262
10	2.228
15	2.131
20	2.086
25	2.060
30	2.042
40	2.021
60	2.000
120	1.980
∞	1.960

Το s_{pooled} είναι μια ομαδοποιημένη τυπική απόκλιση που χρησιμοποιεί και τις 2 σειρές δεδομένων

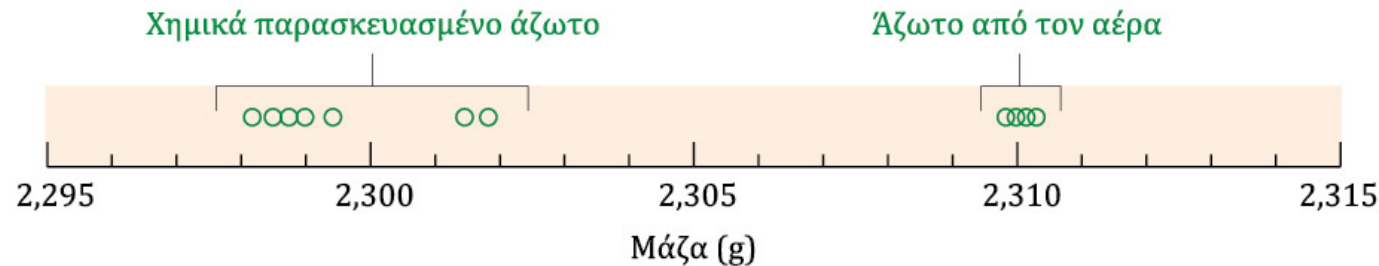
$$t_{\text{calculated}}(0.300) < t_{\text{table}}(2.131) \quad (95\%)$$

- Η διαφορά στις μέσες τιμές δεν είναι σημαντική.
- Αυτό το συμπέρασμα είναι αναμενόμενο καθώς η διαφορά είναι μικρότερη από τις τυπικές αποκλίσεις και των δύο σειρών μετρήσεων.

Περίπτωση 2β: Οι τυπικές αποκλίσεις είναι σημαντικά διαφορετικές

Ο ξηρός αέρας αποτελείται από περίπου ένα πέμπτο οξυγόνο και τέσσερα πέμπτα άζωτο

Ο Rayleigh απομάκρυνε το O_2 από τον αέρα περνώντας τον μέσα από διάλυρο χαλκό προς σχηματισμό CuO



ΕΙΚΟΝΑ 4-7 Οι μετρήσεις του Rayleigh της μάζας σταθερού όγκου αερίου (σε σταθερή πίεση και θερμοκρασία) που είχε απομονωθεί από αέρα ή είχε σχηματιστεί από διάσπαση ενώσεων του αζώτου. Ο Rayleigh αναγνώρισε ότι η διαφορά μεταξύ των δύο σειρών ήταν εκτός της πειραματικής αβεβαιότητας και συμπέρανε ότι ένα βαρύτερο συστατικό, που αποδείχθηκε ότι ήταν το αργόν, ήταν παρόν στο αέριο που είχε απομονωθεί από τον αέρα.

ΜΠ212-Ενόργανη Χημική Ανάλυση-Κεφάλαιο 4

ΠΙΝΑΚΑΣ 4-5 Μάζα αερίου που απομονώθηκε από τον Λόρδο Rayleigh

Από αέρα (g)	Από χημική διάσπαση (g)
2,31017	2,30143
2,30986	2,29890
2,31010	2,29816
2,31001	2,30182
2,31024	2,29869
2,31010	2,29940
2,31028	2,29849
—	2,29889
Μέση τιμή 2,31010 ₉	2,29947 ₂
Τυπική απόκλιση 0.00014 ₃	0,00137 ₉

Πηγή: Δεδομένα από R. D. Larsen, Lessons Learned from Lord Rayleigh on the Importance of Data Analysis, *J. Chem. Ed.* **1990**, 67, 925· επίσης C. J. Giunta, Using History to Teach Scientific Method: The Case of Argon, *J. Chem. Ed.* **1998**, 75, 1322.

Περίπτωση 2β: Οι τυπικές αποκλίσεις είναι σημαντικά διαφορετικές

Δοκιμή F

$$F_{calc} = \frac{s_1^2}{s_2^2} = \frac{(0.00137)^2}{(0.00014)^2} = 93.1$$

$$F_{table} = 5.7$$

Η διαφορά είναι σημαντική

ΠΙΝΑΚΑΣ 4-5 Μάζα αερίου που απομονώθηκε από τον Λόρδο Rayleigh

Από αέρα (g)	Από χημική διάσπαση (g)
2,31017	2,30143
2,30986	2,29890
2,31010	2,29816
2,31001	2,30182
2,31024	2,29869
2,31010	2,29940
2,31028	2,29849
—	2,29889
Μέση τιμή 2,31010 ₉	2,29947 ₂
Τυπική απόκλιση 0.00014 ₃	0,00137 ₉

Πηγή: Δεδομένα από R. D. Larsen, Lessons Learned from Lord Rayleigh on the Importance of Data Analysis, *J. Chem. Ed.* **1990**, 67, 925· επίσης C. J. Giunta, Using History to Teach Scientific Method: The Case of Argon, *J. Chem. Ed.* **1998**, 75, 1322.

Περίπτωση 2β: Οι τυπικές αποκλίσεις είναι σημαντικά διαφορετικές

Όταν οι τυπικές αποκλίσεις δύο σειρών δεδομένων είναι σημαντικά διαφορετικές, οι εξισώσεις για τη δοκιμή t είναι:

$$t_{\text{calculated}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(u_1^2) + (u_2^2)}}$$

Τυπική απόκλιση
της μέσης τιμής

$$u_i = s_i / \sqrt{n_i}$$

$$\text{Βαθμοί ελευθερίας} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = \frac{(u_1^2 + u_2^2)^2}{\frac{u_1^4}{n_1 - 1} + \frac{u_2^4}{n_2 - 1}}$$

Στρογγυλοποιούμε τους βαθμούς ελευθερίας της Εξίσωσης 4-10β στον κοντινότερο ακέραιο.

Περίπτωση 2β: Οι τυπικές αποκλίσεις είναι σημαντικά διαφορετικές

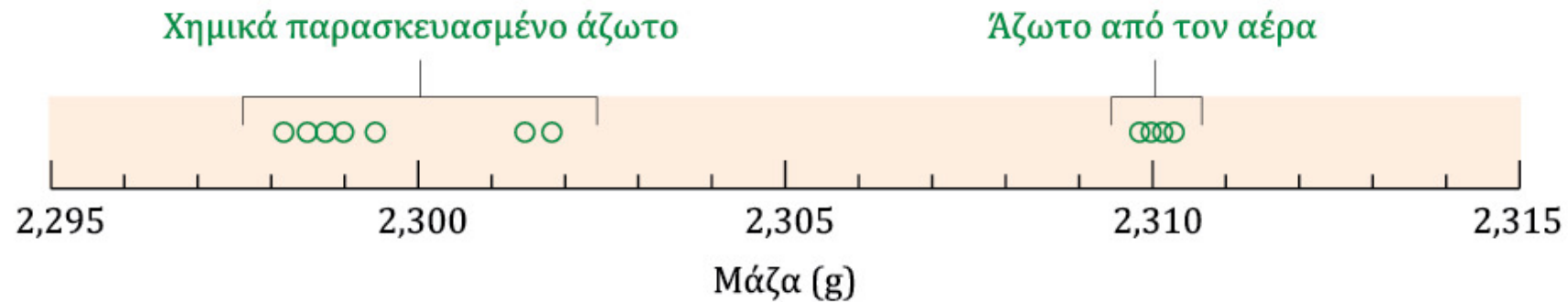
Η δοκιμή F μας απέδειξε ότι οι τυπικές αποκλίσεις είναι σημαντικά διαφορετικές, έτσι χρησιμοποιούμε τις Εξισώσεις 4-9β και 4-10β:

$$t_{\text{calculated}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}} = \frac{|2.310\ 10_9 - 2.299\ 47_2|}{\sqrt{0.000\ 14_3^2/7 + 0.001\ 37_9^2/8}} = 21.7$$

$$\text{Βαθμοί ελευθερίας} = \frac{\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2 + (s_2^2/n_2)^2}}{\frac{n_1-1}{7-1} + \frac{n_2-1}{8-1}} = \frac{(0.000\ 14_3^2/7 + 0.001\ 37_9^2/8)^2}{(0.000\ 14_3^2/7)^2 + (0.001\ 37_9^2/8)^2} = 7.17$$

Περίπτωση 2β: Οι τυπικές αποκλίσεις είναι σημαντικά διαφορετικές

Για 7 βαθμούς ελευθερίας, η κρίσιμη τιμή του t στον Πίνακα 4-4 για στάθμη εμπιστοσύνης 95% είναι 2,365. Η παρατηρούμενη τιμή $t_{\text{calculated}} = 21,7$ υπερβαίνει κατά πολύ την τιμή t_{table} . Η προφανής διαφορά ανάμεσα στις δύο σειρές δεδομένων είναι πολύ σημαντική.



ΕΙΚΟΝΑ 4-7 Οι μετρήσεις του Rayleigh της μάζας σταθερού όγκου αερίου (σε σταθερή πίεση και θερμοκρασία) που είχε απομονωθεί από αέρα ή είχε σχηματιστεί από διάσπαση ενώσεων του αζώτου. Ο Rayleigh αναγνώρισε ότι η διαφορά μεταξύ των δύο σειρών ήταν εκτός της πειραματικής αβεβαιότητας και συμπέρανε ότι ένα βαρύτερο συστατικό, που αποδείχθηκε ότι ήταν το αργόν, ήταν παρόν στο αέριο που είχε απομονωθεί από τον αέρα.

n-1	95%
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306
9	2.262
10	2.228
15	2.131
20	2.086
25	2.060
30	2.042
40	2.021
60	2.000
120	1.980
∞	1.960

Περίπτωση 3. Συζευγμένη (paired) δοκιμή t για σύγκριση συγκεκριμένων διαφορών

Χρησιμοποιούμε δύο μεθόδους για να πραγματοποιήσουμε μεμονωμένες μετρήσεις σε πολλά διαφορετικά δείγματα. Δεν υπάρχουν επαναλήψεις.

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n - 1}}$$

$$t_{\text{calculated}} = \frac{|\bar{d}|}{s_d} \sqrt{n}$$

d: η διαφορά μεταξύ 2 αποτελεσμάτων
d̄: η μέση τιμή των διαφορών

Για 7 βαθμούς ελευθερίας, η κρίσιμη τιμή του t στον Πίνακα 4-4 για στάθμη εμπιστοσύνης 95% είναι 2,365. Η παρατηρούμενη τιμή t_{calculated} = 0.803 πολύ μικρότερη από την τιμή t_{table}. Η διαφορά ανάμεσα στις δύο σειρές δεδομένων δεν είναι σημαντική.

	A	B	C	D
1	Comparison of methods for measuring nitrate			
2				
3		Nitrate (ppm) in plant extract		
4		Spectrophotometry	Experimental	
5	Sample	with Cd reduction	biosensor	Difference (d _i)
6	1	1,22	1,23	0,01
7	2	1,21	1,58	0,37
8	3	4,18	4,04	-0,14
9	4	3,96	4,92	0,96
10	5	1,18	0,96	-0,22
11	6	3,65	3,37	-0,28
12	7	4,36	4,48	0,12
13	8	1,61	1,70	0,09
14			Mean difference =	0,114
15			Standard deviation of differences =	0,401
16			Number of samples =	8
17	D6 = C6-B6		t _{calculated} =	0,803
18	D14 = AVERAGE(D6:D13)		t _{table} =	2,365
19	D15 = STDEV.S(D6:D13)			
20	D16 = COUNT(D6:D13)			
21	D17 = ABS(D14)*SQRT(A13)/D15 (ABS = absolute value)			
22	D18 = T.INV.2T(0,05,D16-1)			

Η Εικόνα 4-8 χρησιμοποιεί τη συνάρτηση του Excel = T.INV.2T(0,05,D16-1) για να υπολογίσει το t_{table} = 2,365 στο κελί D18 για (1 - 0,05) = 95% στάθμη εμπιστοσύνης και 8 - 1 = 7 βαθμούς ελευθερίας.

ΕΙΚΟΝΑ 4-8 Μετρήσεις νιτρικών σε εκχυλίσματα φυτών με δύο μεθόδους. [Δεδομένα από N. Plumeré, J. Henig, and W. H. Campbell, "Enzyme-Catalyzed O2 Removal System for Electrochemical Analysis Under Ambient Air: Application in an Amperometric Nitrate Biosensor," *Anal. Chem.* **2012**, 84, 2141].

Μονοπλευρικές (one-tailed) και διπλευρικές (two-tailed) δοκιμές σημαντικότητας

Αρσενικό/L σε δείγματα πόσιμου νερού= 10 µg As/L

Δείγματα νερού→ 10.06, 10.12, 10.19, and 10.04 µg As/L

- Σύγκριση μέσης τιμής 4 μετρήσεων με πιστοποιημένη τιμή

ΔΙΠΛΕΥΡΗ: Εφόσον η κατανομή t είναι συμμετρική (2.5% του εμβαδού βρίσκεται πάνω από το $t = 3.182$ και 2.5% του εμβαδού βρίσκεται κάτω από το $t = - 3.182$), η τιμή που ψάχνουμε πρέπει να είναι η δίπλευρη τιμή του $t = 3.182$ για στάθμη εμπιστοσύνης 95% στον Πίνακα 4-4.

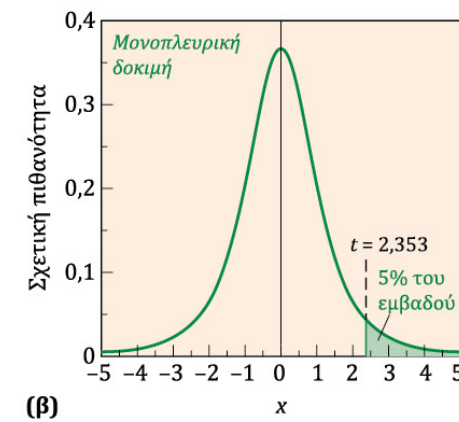
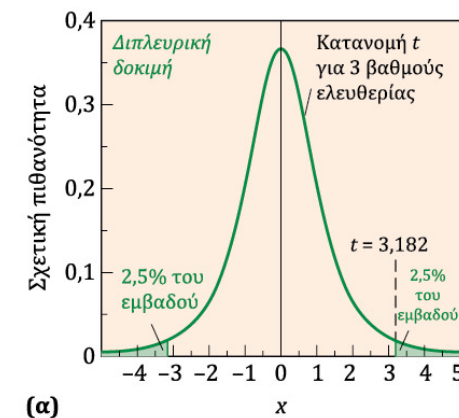
ΜΟΝΟΠΛΕΥΡΗ: Μας ενδιαφέρει μόνο η πιθανότητα το αρσενικό να υπερβαίνει το όριο. Η Εικόνα 4- 9β δείχνει ότι το 5% του εμβαδού κάτω από την καμπύλη βρίσκεται πάνω από το $t = 2,353$

$$\bar{x} = 10.10_{25} \mu\text{g/L} \quad s=0.06_{75} \mu\text{g/L}$$

$$t_{\text{calculated}} = \frac{|\bar{x} - \text{regulatory limit}|}{s} \sqrt{n} = \frac{|10.10_{25} - 10|}{0.06_{75}} \sqrt{4} = 3.0_4$$

$$t_{\text{calculated}}(3.04) > t_{\text{table}}(2.353) \quad (95\%)$$

Άρα μηδενική υπόθεση απορρίπτεται και το δείγμα περιέχει περισσότερο αρσενικό από το επιτρεπτό.



ΕΙΚΟΝΑ 4-9 Κατανομή Student's t για 3 βαθμούς ελευθερίας. Στο πλαίσιο α, κάθε σκιασμένη περιοχή περιέχει το 2,5% του εμβαδού κάτω από την καμπύλη. Στο πλαίσιο β, η μοναδική σκιασμένη περιοχή περιέχει το 5% του εμβαδού. Όσο λιγότεροι είναι οι βαθμοί ελευθερίας, τόσο ευρύτερη είναι η κατανομή. Καθώς αυξάνονται οι βαθμοί ελευθερίας, η καμπύλη πλησιάζει την καμπύλη Gauss.

n-1	90%	95%
1	6.314	12.706
2	2.920	4.303
3	2.353	3.182
4	2.132	2.776
5	2.015	2.571
6	1.943	2.447
7	1.895	2.365
8	1.860	2.306
9	1.833	2.262
10	1.812	2.228
15	1.753	2.131
20	1.725	2.086
25	1.708	2.060
30	1.697	2.042
40	1.684	2.021
60	1.671	2.000
120	1.658	1.980
∞	1.645	1.960

Μονοπλευρικές (one-tailed) και διπλευρικές (two-tailed) δοκιμές σημαντικότητας

Μετρήσεις δείγματος ελέγχου	Συγκ. δείγματος ελέγχου
90 mg/dL	99 mg/dL
94 mg/dL	
85 mg/dL	
95 mg/dL	
93 mg/dL	
Μέση τιμή = 91.4 mg/dL	



n-1	98	99
1	31.821	63.656
2	6.965	9.925
3	4.541	5.841
4	3.747	4.604
5	3.365	4.032
6	3.143	3.707
7	2.998	3.500
8	2.896	3.355
9	2.821	3.250
10	2.764	3.169
15	2.602	2.947
20	2.528	2.845
25	2.485	2.787
30	2.457	2.750
40		
60		
120		
∞		

$$t_{\text{calculated}} = \frac{|\bar{x} - \text{πιστοπ. τιμή}|}{s}$$

$$\sqrt{n} = \frac{|91.4 - 99|}{4.0} \sqrt{5} = 4.25$$

$$3.747 \text{ (98\%)} < t_{\text{cal}} < 4.604 \text{ (99\%)}$$

Πιθανότητα <2% για να συμφωνούν μετρήσεις με αυτή του κατασκευαστή
 Πρόβλημά με συσκευή ή άτομο που έκανε τη μέτρηση

4-5 Δοκιμές t με υπολογιστικό φύλλο

Δοκιμές t με υπολογιστικό φύλλο

Υπολογιστικό φύλλο για τη σύγκριση των μέσων τιμών των μετρήσεων αζώτου του Rayleigh

Δεδομένα σε αρχείο στο macmillanlearning.com

	A	B	C	D	E	F	G
1	Analysis of Rayleigh's Data				t-Test: Two-Sample Assuming Equal Variances		
2						Variable 1	Variable 2
3		Mass of gas (g) collected from			Mean	2,310109	2.299473
4		air	chemical		Variance	2,03E-08	1.9E-06
5		2,31017	2,30143		Observations	7	8
6		2,30986	2,29890		Pooled Variance	1,03E-06	
7		2,31010	2,29816		Hypothesized Mean Diff	0	
8		2,31001	2,30182		df	13	
9		2,31024	2,29869		t Stat	20,21372	
10		2,31010	2,29940		P(T<=t) one-tail	1,66E-11	
11		2,31028	2,29849		t Critical one-tail	1,770932	
12			2,29889		P(T<=t) two-tail	3,32E-11	
13	Average	2,31011	2,29947		t Critical two-tail	2,160368	
14	Std Dev	0,00014	0,00138				
15					t-Test: Two-Sample Assuming Unequal Variances		
16	B13 = AVERAGE(B5:B12)					Variable 1	Variable 2
17	B14 = STDEV.S(B5:B12)				Mean	2,310109	2,299473
18					Variance	2,03E-08	1,9E-06
19					Observations	7	8
20					Hypothesized Mean Diff	0	
21					df	7	
22					t Stat	21,68022	
23					P(T<=t) one-tail	5,6E-08	
24					t Critical one-tail	1,894578	
25					P(T<=t) two-tail	1,12E-07	
26					t Critical two-tail	2,364623	

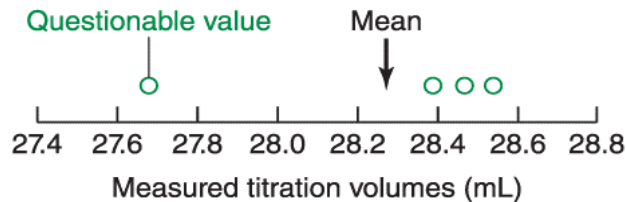
ΕΙΚΟΝΑ 4-10 Υπολογιστικό φύλλο για να συγκρίνουμε τις μέσες τιμές των μετρήσεων του Rayleigh's του Πίνακα 4-5.
ΜΠ212-Ενόργανη Χημική Ανάλυση-Κεφάλαιο 4

4-6 Δοκιμή Grubbs για αποκλίνουσες τιμές

Δοκιμή Grubbs για αποκλίνουσες τιμές

$$G_{\text{calculated}} = \frac{|\text{αμφισβητούμενη τιμή} - \bar{x}|}{s}$$

Αν το $G_{\text{calculated}} > G_{\text{table}}$, η ύποπτη τιμή μπορεί να απορριφθεί με εμπιστοσύνη 95%. Οι τιμές σε αυτόν τον Πίνακα είναι για μονόπλευρη δοκιμή, όπως προτείνεται από τον ASTM.



Harris/Lucy, *Quantitative Chemical Analysis*,
10e, © 2020 W. H. Freeman and Company

Αριθμός παρατηρήσεων	G (95% εμπιστοσύνη)	Αριθμός παρατηρήσεων	G (95% εμπιστοσύνη)
3	1.153	10	2.176
4	1.463	11	2.234
5	1.672	12	2.285
6	1.822	15	2.409
7	1.938	20	2.557
8	2.032	30	2.745
9	2.110	50	2.956

Δοκιμή Grubbs για αποκλίνουσες τιμές

Όγκος τιτλοδότησης προτύπου(mL): **28.54, 28.39, 28.47, 27.68**

$\bar{x} = 28.27 \pm 0.40 \text{ mL}$; $\text{RSD} = 1.4\%$ ← (τιμή μεγαλύτερη από την αναμενόμενη ακρίβεια)

$$G_{\text{calculated}} = \frac{|27.68 - 28.27|}{0.40} = 1.482 \quad G_{\text{table}} = 1.463 \text{ για τις τέσσερις παρατηρήσεις}$$

$G_{\text{calculated}} > G_{\text{table}}$, το αμφισβητούμενο σημείο θα πρέπει να απορριφθεί.

4-7 Η μέθοδος των ελαχίστων τετραγώνων

Η μέθοδος των ελαχίστων τετραγώνων

Στις περισσότερες αναλύσεις, η απόκριση αξιολογείται με γνωστές συγκεντρώσεις του αναλύτη (τα πρότυπα) έτσι ώστε να μπορεί να αξιοποιηθεί η απόκριση ενός άγνωστου δείγματος. Τις περισσότερες φορές εργαζόμαστε στην περιοχή που η καμπύλη βαθμονόμησης είναι ευθεία.

Χρησιμοποιούμε τη **μέθοδο των ελαχίστων τετραγώνων** για να **σχεδιάσουμε** την «καλύτερη» ευθεία που περνάει από πειραματικά σημεία που έχουν κάποια διασπορά

- Κάποια σημεία θα βρίσκονται πάνω από τη γραμμή και κάποια κάτω από τη γραμμή.
- Εξίσωση ευθείας $y = mx + b$

Εύρεση της εξίσωσης της ευθείας

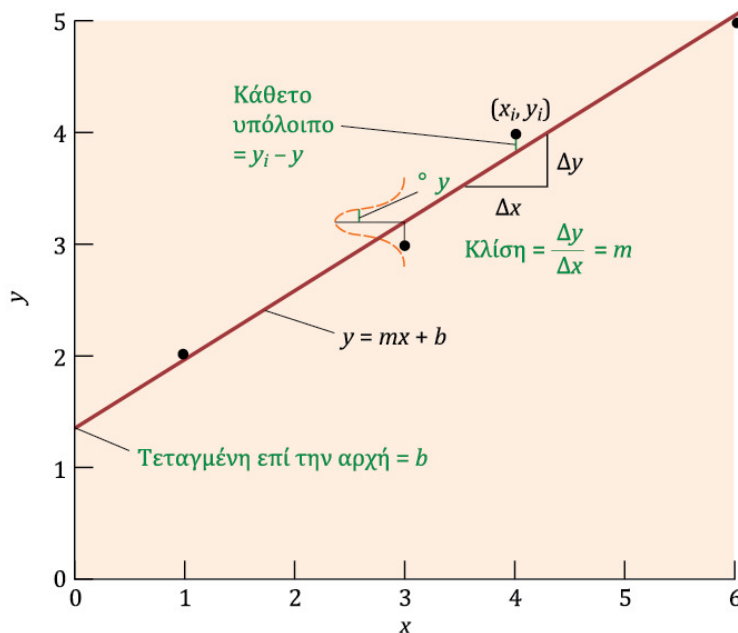
Θέλουμε να σχεδιάσουμε την βέλτιστη ευθεία γραμμή με τα σημεία ελαχιστοποιώντας τα κάθετα υπόλοιπα ανάμεσα στα σημεία και τη γραμμή

$$y = mx + b$$

Slope y-intercept

Κάθετο υπόλοιπο $d_i = y_i - y = y_i - (mx_i + b)$

Κάποια υπόλοιπα είναι θετικά και κάποια είναι αρνητικά. Εφόσον θέλουμε να ελαχιστοποιήσουμε την τιμή των υπολοίπων ανεξάρτητα από το πρόσημό τους, υψώνουμε όλα τα υπόλοιπα στο τετράγωνο



Κάθετο υπόλοιπο $= y_i - y$

$$d_i^2 = (y_i - mx_i + b)^2 \leftarrow \text{μέθοδος ελαχίστων τετραγώνων}$$

ΠΑΡΑΔΕΙΓΜΑ: Εύρεση της κλίσης και της τεταγμένης επί την αρχή με υπολογιστικό φύλλο

Το Excel έχει συναρτήσεις που ονομάζονται SLOPE, INTERCEPT και LINEST

	A	B	C	D	E	F
1	x	y			Formulas:	
2	1	2		slope =		
3	3	3		0.61538	D3 = SLOPE(B2:B5,A2:A5)	
4	4	4		intercept =		
5	6	5		1.34615	D5 = INTERCEPT(B2:B5,A2:A5)	

Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

	A	B	C	D	E	F	G
1	x	y			Output from LINEST		
2	1	2			Slope	Intercept	
3	3	3		Parameter	0.61538	1.34615	
4	4	4		u_m	0.05439	0.21414	u_b
5	6	5		R^2	0.98462	0.19612	s_y
6	Highlight cells E3:F5						
7	Type "= LINEST(B2:B5,A2:A5,TRUE,TRUE)"						
8	Press CTRL+SHIFT+ENTER (on PC)						
9	Press CTRL+SHIFT+RETURN (on Mac)						

Το πρώτο ψηφίο αβεβαιότητας είναι το τελευταίο σημαντικό ψηφίο. Συχνά διατηρούμε παραπάνω ψηφία για να αποφύγουμε σφάλματα στρογγυλοποίησης σε υπολογισμούς που ακολουθούν.

Κλίση:

$0,61538 \pm 0,05439 = 0,62 \pm 0,05 \text{ ή } 0,61_5 \pm 0,05_4$

(4-23)

Τομή με τον άξονα των y:

$1,34615 \pm 0,21415 = 1,3 \pm 0,2 \text{ ή } 1,3_5 \pm 0,2_1$

(4-24)

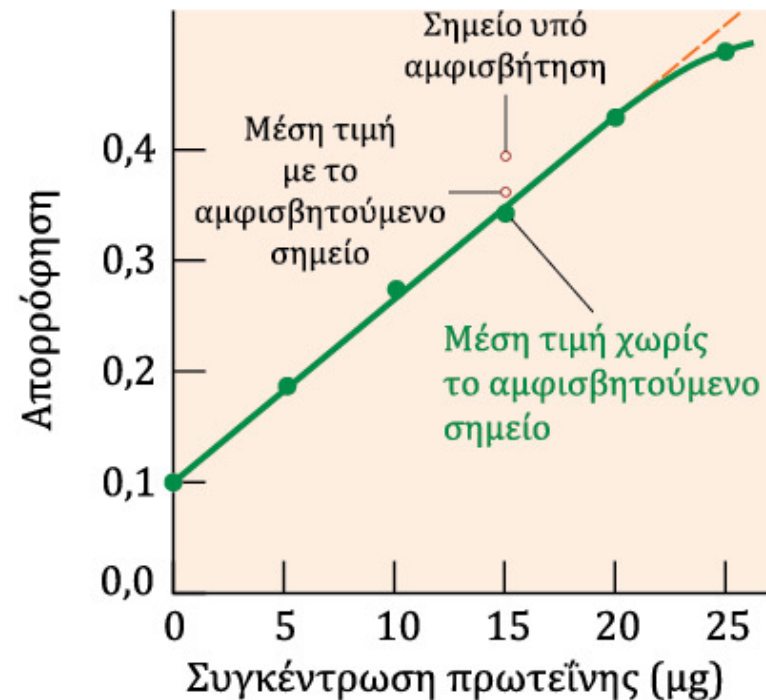
4-8 Καμπύλες βαθμονόμησης

Καμπύλες βαθμονόμησης

Μια καμπύλη βαθμονόμησης καταγράφει την απόκριση μιας αναλυτικής μεθόδου συναρτήσει γνωστών ποσοτήτων αναλύτη.

- Διαλύματα που περιέχουν γνωστές συγκεντρώσεις αναλύτη ονομάζονται **πρότυπα διαλύματα**.
- Τα διαλύματα που περιέχουν τους διαλύτες και τα αντιδραστήρια της ανάλυσης, αλλά όχι αναλύτη, ονομάζονται **λευκά διαλύματα**.

Ένα φασματοφωτόμετρο μετράει την απορρόφηση του φωτός, που είναι ανάλογη της ποσότητας της πρωτεΐνης που υπάρχει στο δείγμα.



ΕΙΚΟΝΑ 4-12 Μέσες τιμές απορρόφησης του Πίνακα 4-8 συναρτήσει των μικρογραμμαρίων της πρωτεΐνης. Οι μέσες τιμές στην περιοχή 0 έως 20 μg πρωτεΐνης βρίσκονται σε μια ευθεία αν παραλείψουμε το αμοιβαίο σημείο 0,392 στα 15 μg.

ΠΙΝΑΚΑΣ 4-8 Φασματοφωτομετρικά δεδομένα για κατασκευή καμπύλης βαθμονόμησης

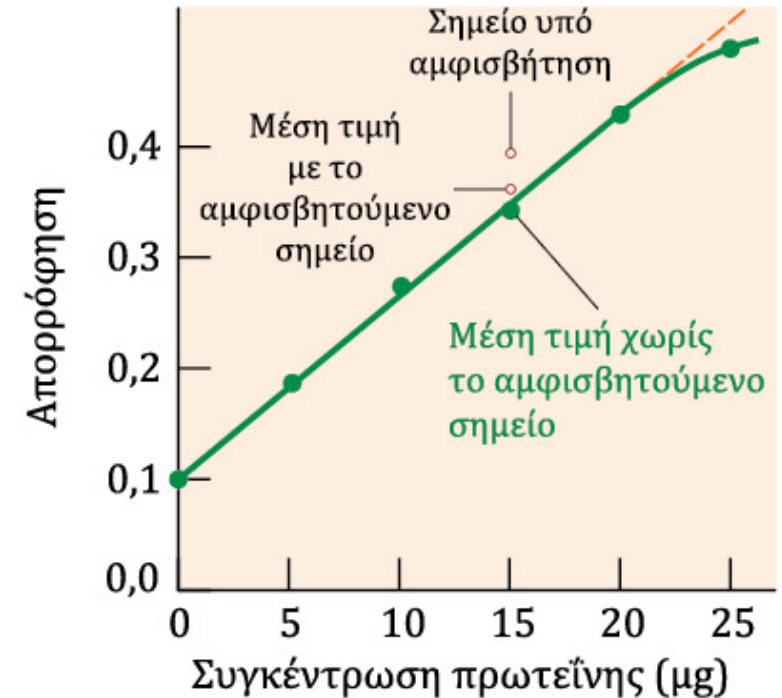
Ποσότητα πρωτεΐνης (μg)	Απορρόφηση ανεξάρτητων προτύπων			Εύρος	Διορθωμένη απορρόφηση		
0	0.099	0.099	0.100	0.001	-0.000 ₃	-0.000 ₃	0.000 ₇
5.0	0.185	0.187	0.188	0.003	0.085 ₇	0.087 ₇	0.088 ₇
10.0	0.282	0.272	0.272	0.010	0.182 ₇	0.172 ₇	0.172 ₇
15.0	0.345	0.347	(0.392)	0.047	0.245 ₇	0.247 ₇	—
20.0	0.425	0.425	0.430	0.005	0.325 ₇	0.325 ₇	0.330 ₇
25.0	0.483	0.488	0.496	0.013	0.383 ₇	0.388 ₇	0.396 ₇

Κατασκευή καμπύλης βαθμονόμησης

1. Παρασκευάζουμε γνωστά διαλύματα αναλύτη που να καλύπτουν το εύρος (0-150%) των αναμενόμενων συγκεντρώσεων του αναλύτη.
2. Αφαιρούμε τη μέση απορρόφηση (0,0993) των λευκών διαλυμάτων από κάθε τιμή απορρόφησης για να υπολογιστεί η διορθωμένη απορρόφηση
3. Κατασκευάζουμε καμπύλη της διορθωμένης απορρόφησης συναρτήσει της ποσότητας της πρωτεΐνης σε κάθε πρότυπο
4. Χρησιμοποιούμε τη μέθοδο ελαχίστων τετραγώνων για να προσαρμόσουμε τη βέλτιστη ευθεία στο γραμμικό εύρος των δεδομένων.
5. Αν αργότερα αναλύσουμε ένα άγνωστο, αναλύουμε ταυτόχρονα ένα λευκό.

$$y = m \times (\mu\text{g πρωτεΐνης}) + b = 0.0163_0 \times (\mu\text{g πρωτεΐνης}) + 0.004_7$$

ΜΠ212-Ενόργανη Χημική Ανάλυση-Κεφάλαιο 4



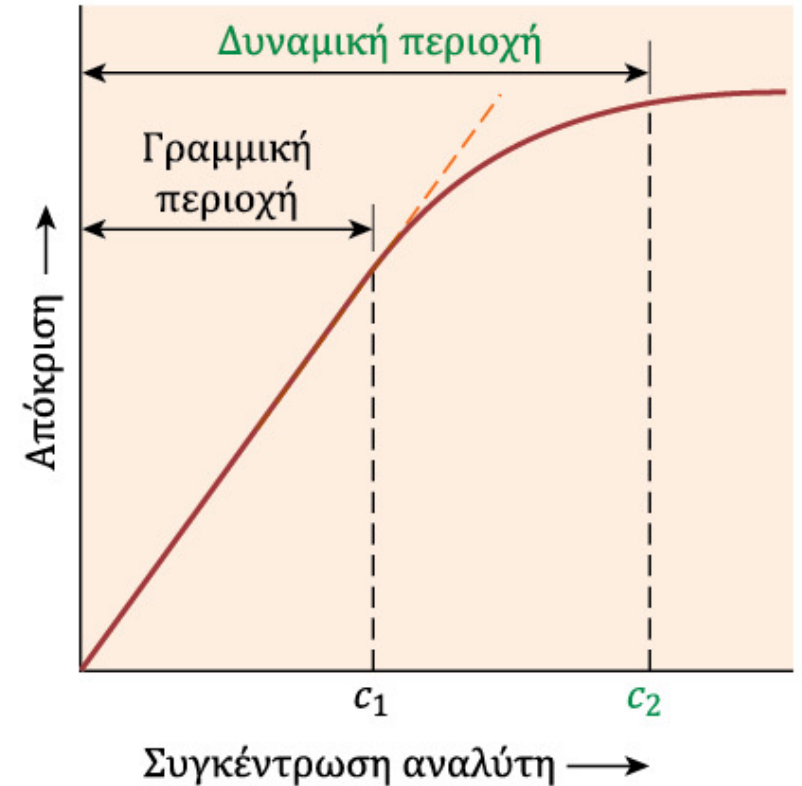
EIKONA 4-12 Μέσες τιμές απορρόφησης του Πίνακα 4-8 συναρτήσει των μικρογραμμαρίων της πρωτεΐνης. Οι μέσες τιμές στην περιοχή 0 έως 20 μg πρωτεΐνης βρίσκονται σε μια ευθεία αν παραλείψουμε το αμφισβητούμενο σημείο 0,392 στα 15 μg.

Γραμμική απόκριση

Η γραμμική περιοχή μιας αναλυτικής μεθόδου είναι το εύρος συγκεντρώσεων του αναλύτη στο οποίο η απόκριση είναι ανάλογη με τη συγκέντρωση.

Μια σχετική παράμετρος στην Εικόνα 4-14 είναι η **δυναμική περιοχή** – η περιοχή συγκεντρώσεων στην οποία υπάρχει μια μετρήσιμη απόκριση ως προς τον αναλύτη έστω κι αν η απόκριση δεν είναι γραμμική.

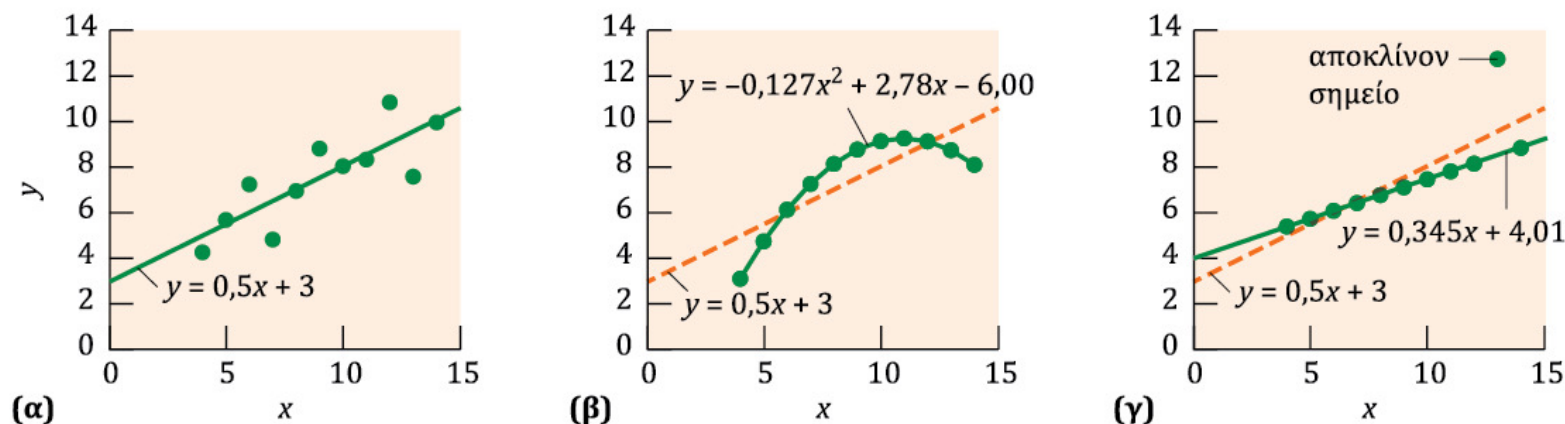
- Προτιμούμε διαδικασίες βαθμονόμησης με γραμμική απόκριση.
- Το διορθωμένο αναλυτικό σήμα (= σήμα από το δείγμα – σήμα από το λευκό) είναι ανάλογο με την ποσότητα του αναλύτη.
- Παρόλο που προσπαθούμε να εργαζόμαστε στη γραμμική περιοχή, μπορούμε να πάρουμε σωστά αποτελέσματα και εκτός της γραμμικής περιοχής.



ΕΙΚΟΝΑ 4-14 Καμπύλη βαθμονόμησης όπου παρουσιάζονται η γραμμική και η δυναμική περιοχή.

Καλές πρακτικές

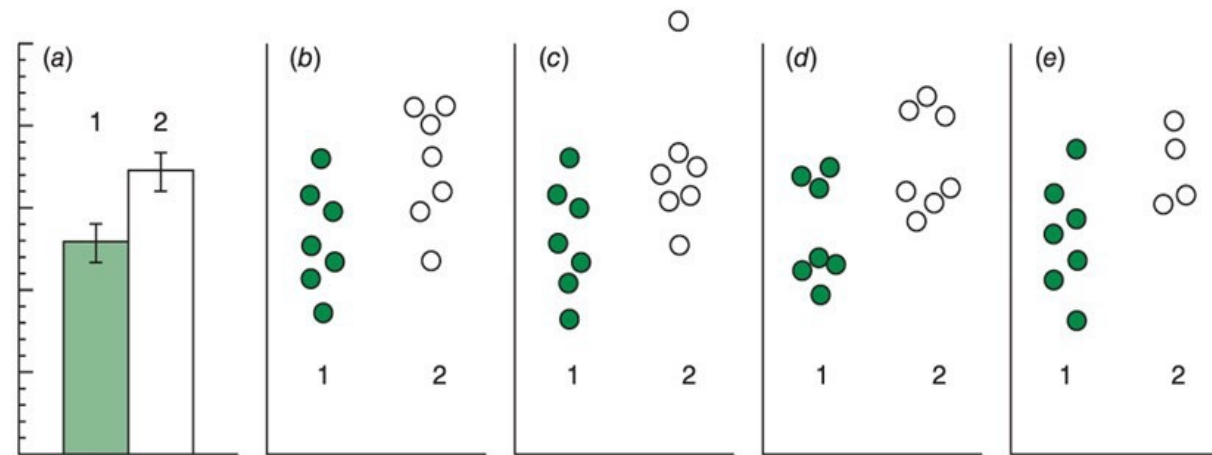
- Πάντα κατασκευάζουμε γράφημα των δεδομένων μας
- Το γράφημα μας δίνει τη δυνατότητα να απορρίπτουμε «κακά» δεδομένα, μας καθοδηγεί να επαναλάβουμε μια μέτρηση και μας βοηθά να αποφασίσουμε αν μια ευθεία είναι κατάλληλη
- Και οι τρεις σειρές δεδομένων στην Εικόνα 4-15 προσαρμόζονται στην ευθεία $y = 0.5x + 3$



ΕΙΚΟΝΑ 4-15 Διαφορετικές κατανομές δεδομένων που προσαρμόζονται στην ευθεία $y = 0.5x + 3$ με τη μέθοδο ελαχίστων τετραγώνων. [Δεδομένα από F. J. Anscombe, "Graphs in Statistical Analysis," *Am. Statistician* **1973**, 27, 17].

ΠΛΑΙΣΙΟ 4-3 Η σημασία των γραφημάτων για την οπτικοποίηση των δεδομένων

Μια καλή γραφική παράσταση αποκαλύπτει σημαντικά χαρακτηριστικά των δεδομένων και καθοδηγεί τη στατιστική τους ανάλυση.



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

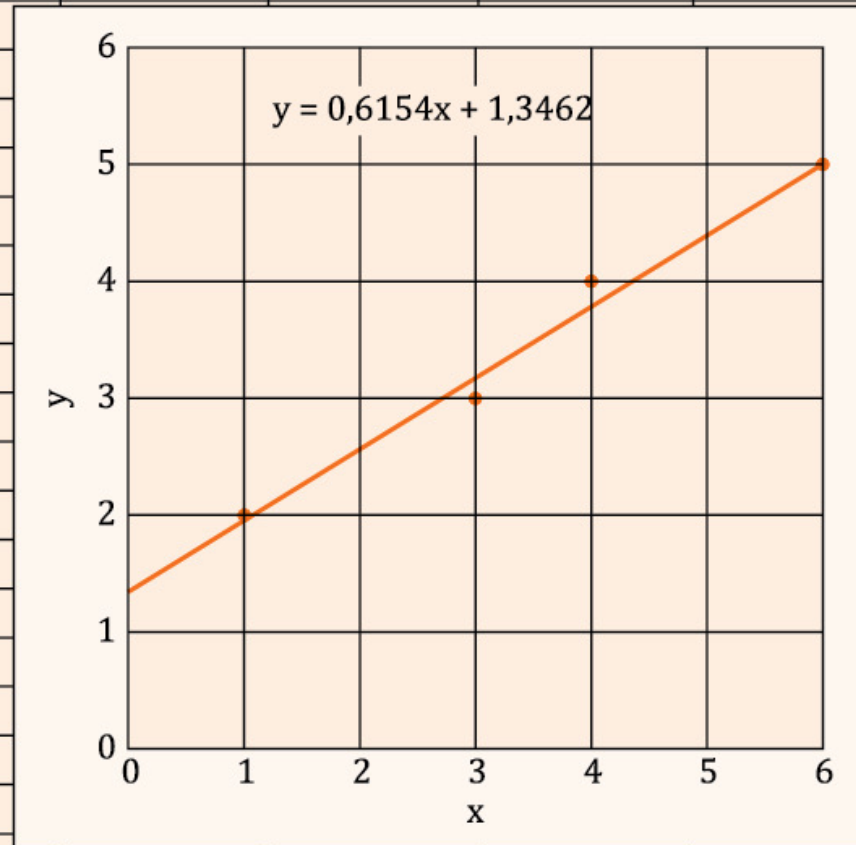
- Τα ύψη στα ραβδογράμματα του πλαισίου (a) παρουσιάζουν τους μέσους όρους δύο σειρών δεδομένων. Οι γραμμές σφάλματος αντιστοιχούν σε τυπική απόκλιση (τυπική απόκλιση της μέσης τιμής).
- Τα σημειογράμματα αναδεικνύουν διαφορετικά χαρακτηριστικά των δεδομένων που δεν είναι εμφανή στο ραβδόγραμμα.

Καλές πρακτικές

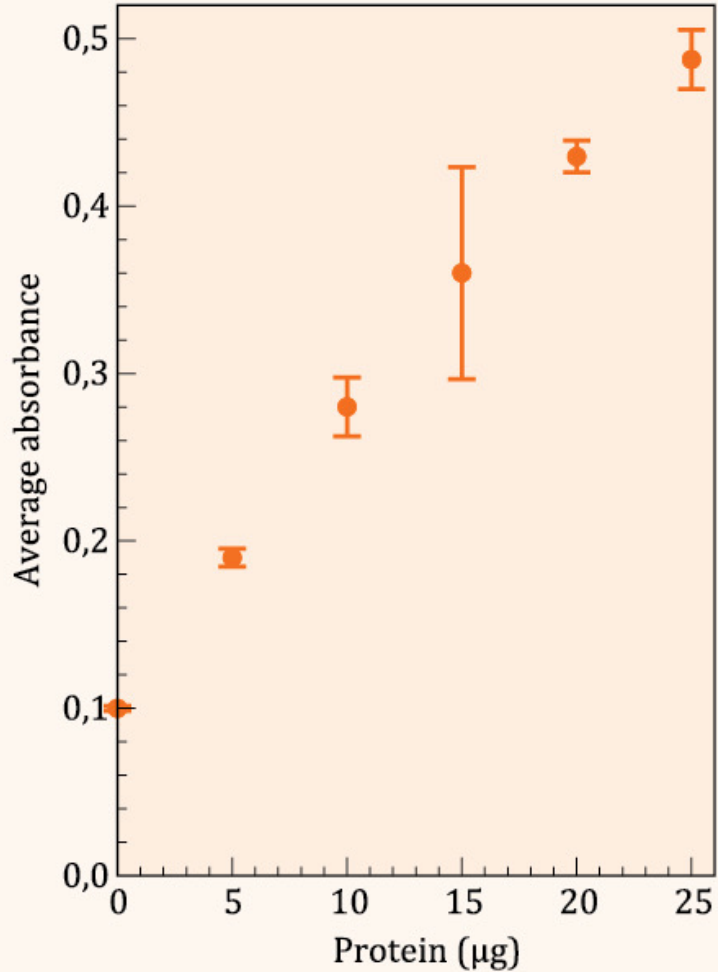
- Δεν είναι αξιόπιστο να επεκτείνουμε τις καμπύλες βαθμονόμησης, είτε είναι γραμμικές είτε όχι, πέρα από το μετρούμενο εύρος προτύπων.
- Προτείνονται το λιγότερο έξι συγκεντρώσεις βαθμονόμησης και δύο επαναλήψεις για κάθε άγνωστο.
- Η πιο αυστηρή προσέγγιση είναι να παρασκευάζουμε ανεξάρτητα κάθε συγκέντρωση βαθμονόμησης από πιστοποιημένο υλικό.
- Αποφεύγουμε τη σειριακή αραίωση από διάλυμα παρακαταθήκης γιατί η σειριακή αραίωση μεταφέρει το συστηματικό σφάλμα που τυχόν υπάρχει στο διάλυμα παρακαταθήκης.
- Μετράμε τα διαλύματα βαθμονόμησης με τυχαία σειρά και όχι με σειρά αυξανόμενης συγκέντρωσης.

4-9 Υπολογιστικό φύλλο για τη μέθοδο ελαχίστων τετραγώνων

	A	B	C	D	E	F	G	H	I
1	Least-Squares Spreadsheet								
2									
3	Highlight cells B10:C12	x	y						
4	Type "= LINEST(C4:C7,	1	2						
5	B4:B7,TRUE,TRUE)	3	3						
6	For PC, press	4	4						
7	CTRL+SHIFT+ENTER	6	5						
8	For Mac, press								
9	CTRL+SHIFT+RETURN	LINEST output:							
10	m	0,6154	1,3462	b					
11	u_m	0,0544	0,2141	u_b					
12	R^2	0,9846	0,1961	s_y					
13									
14	n =	4	B14 = COUNT(B4:B7)						
15	Mean y =	3,5	B15 = AVERAGE(C4:C7)						
16	$\Sigma(x_i - \text{mean } x)^2 =$	13	B16 = DEVSQ(B4:B7)						
17									
18	Measured y =	2,72	Input						
19	k = Number of replicate measurements of y =	1	Input						
20	Derived x =	2,2325	B20 = (B18-C10)/B10						
21	$u_x =$	0,3735	B21 = (C12/ABS(B10))*SQRT((1/B19)+(1/B14)+((B18-B15)^2)/(B10^2*B16))						



EIKONA 4-16 Υπολογιστικό φύλλο για ανάλυση με τη μέθοδο ελαχίστων τετραγώνων.

	A	B	C	D	E	F	G	H
1	Adding error bars to a graph							
2	Protein	Mean	Std dev	95%				
3	(μg)	absorbance	absorbance	confidence				
4	0	0,0993	0,00058	0,0014				
5	5	0,1867	0,00153	0,0038				
6	10	0,2753	0,00577	0,0143				
7	15	0,3613	0,02658	0,0660				
8	20	0,4267	0,00289	0,0072				
9	25	0,4890	0,00656	0,0163				
10								
11	95% confidence: D4 = T.INV.2T(0,05,2)*C4/SQRT(3)							
12								
13	Excel 2016 instructions:							
14	To insert error bars for uncertainties in column D:							
15	Double click in graph to get Chart Tools/Design ribbon							
16	Click Add Chart Element, select Error Bars							
17	Choose More Error Bar Options							
18	In Error Bar Options menu, click bar graph icon							
19	Click Custom and Specify Value							
20	For Positive Error Value, highlight D4:D9							
21	For Negative Error Value, highlight D4:D9							
22	Click OK and error bars appear on the chart							
23	Click on one x error bar and press Delete to remove all x error bars.							

EIKONA 4-17 Προσθήκη γραμμών σφάλματος 95% σε γράφημα.