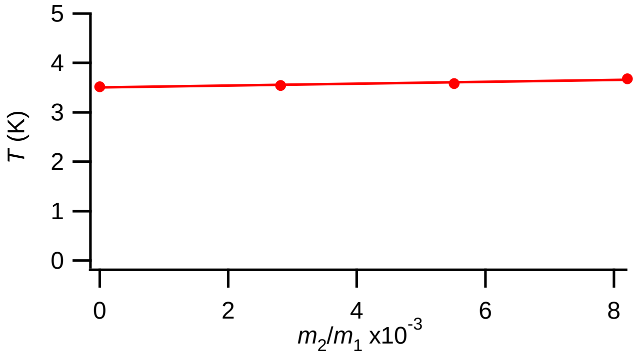
Ορθοί και λανθασμένοι τρόποι απεικονίσεως δεδομένων σε διάγραμμα

Από μετρήσεις σημείου ζέσεως σειράς διαλυμάτων προκύπτουν τα εξής δεδομένα:

***T* (K)**	**m_2/m_1**
3.52	0.00000
3.54	0.00281
3.58	0.00552
3.68	0.00821

Σύμφωνα με την θεωρία τα δεδομένα πρέπει να περιγράφονται από ευθεία με εξίσωση $y = a\,x + b$, όπου $y = T$ και $x = m_2/m_1$.

Κατασκευάζουμε διάγραμμα με τις τιμές αυτές:

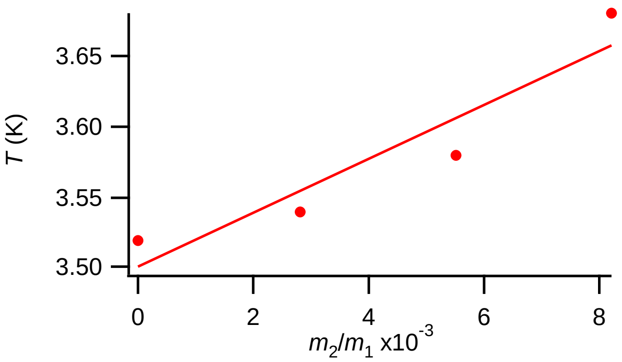


Υπολογίζουμε την εξίσωση της ευθείας με την μέθοδο ελαχίστων τετραγώνων και βρίσκουμε ότι $y = 18.97\,x + 3.502$. Η στατιστική επεξεργασία προσδιορίζει και τις αβεβαιότητες των a και b: $\sigma_a = 4.9$, $\sigma_b = 0.025$. Πώς εξηγείται η μεγάλη τιμή του σ_a η οποία φαίνεται παράδοξη κρίνοντας από την ποιότητα του διαγράμματος με το μάτι; Μήπως το διάγραμμα αντί να αναδεικνύει την καλή ποιότητα των μετρήσεων συγκαλύπτει την κακή τους ποιότητα με ατυχή επιλογή των διαστημάτων τιμών στους άξονες;

Κατασκευάζουμε το ίδιο διάγραμμα έχοντας αναπτύξει τις τιμές του κατακόρυφου άξονα ώστε να αξιοποιείται όλος ο διαθέσιμος χώρος από τις μετρήσεις.

Δεν χρειάζεται να επαναλάβουμε τον υπολογισμό της μεθόδου των ελαχίστων τετραγώνων γιατί τα δεδομένα είναι τα ίδια, άρα και η εξίσωση της ευθείας είναι η ίδια με πριν. Όμως τώρα είναι σαφές γιατί έχει μεγάλη αβεβαιότητα η κλίση της ευθείας.

Δηλ. για να φανεί η ποιότητα των μετρήσεων και η διακύμανση των τιμών είναι απαραίτητο να γίνεται τέτοια επιλογή του διαστήματος τιμών στους άξονες ώστε να αξιοποιείται όλη η επιφάνεια του διαγράμματος.

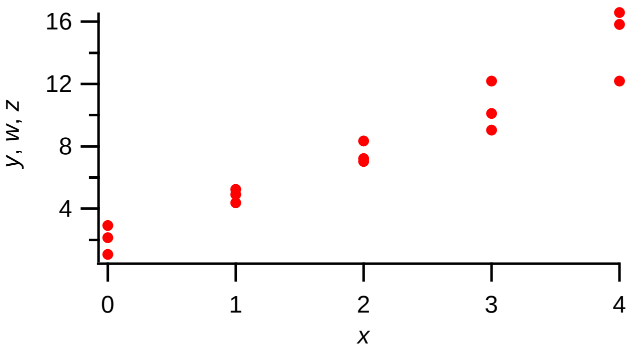


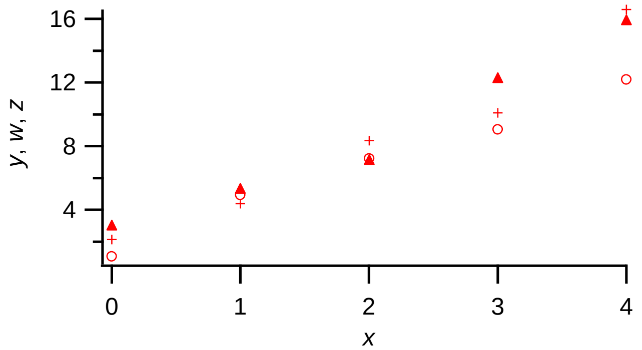
Όταν σε ένα διάγραμμα θέλουμε να παραστήσουμε 2 ή περισσότερες σειρές μετρήσεων, είναι σκόπιμο να χρησιμοποιούμε κατάλληλο συμβολισμό για να διακρίνουμε ποια σημεία αντιστοιχούν σε ποια σειρά μετρήσεων. Για αυτό το λόγο ποικίλουμε το χρώμα ή το σύμβολο (+, *, □, ■, ●, ○) και σημειώνουμε σε υπόμνημα τι παριστάνουμε με κάθε σύμβολο και χρώμα.

Δίνονται 3 σειρές μετρήσεων:

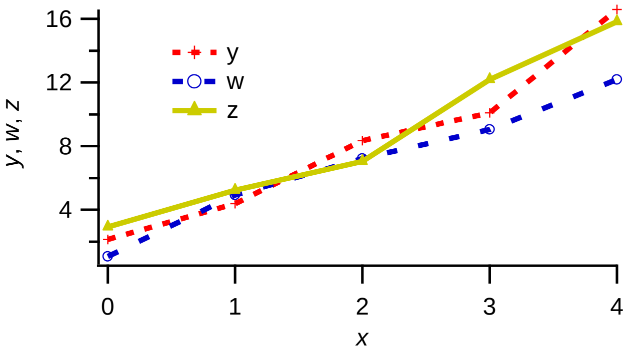
x	*y*	*w*	*z*
0.00	2.13	1.07	2.92
1.00	4.39	4.92	5.24
2.00	8.34	7.22	7.04
3.00	10.10	9.05	12.19
4.00	16.58	12.18	15.81

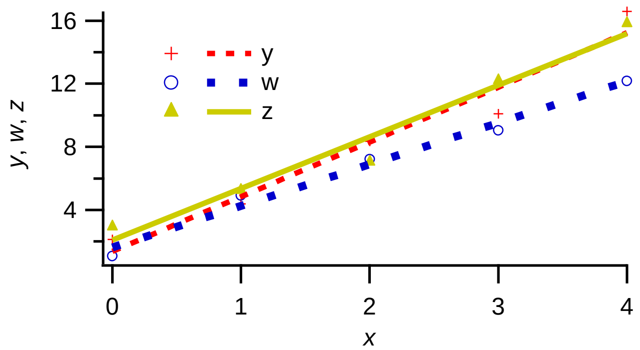
Σχεδιάζουμε διάγραμμα με όλες τις τιμές *y*, *w*, *z* συναρτήσει των τιμών *x*, αλλά δεν μπορούμε να διακρίνουμε πώς ομαδοποιούνται οι τιμές. Όμως αν αλλάξουμε συμβολισμό, αυτό είναι εφικτό:





ή, ακόμη καλύτερα:





Αν εφαρμόσουμε την μέθοδο ελαχίστων τετραγώνων στις 3 σειρές μετρήσεων προκύπτουν οι ευθείες με εξισώσεις αντίστοιχα:

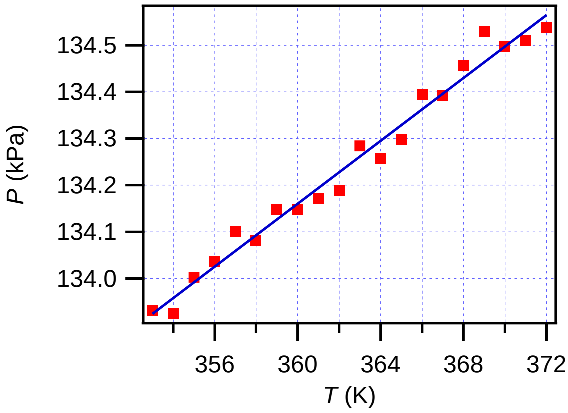
$y = 3.46\ (\pm 0.4)\ x + 1.4\ (\pm 1)$

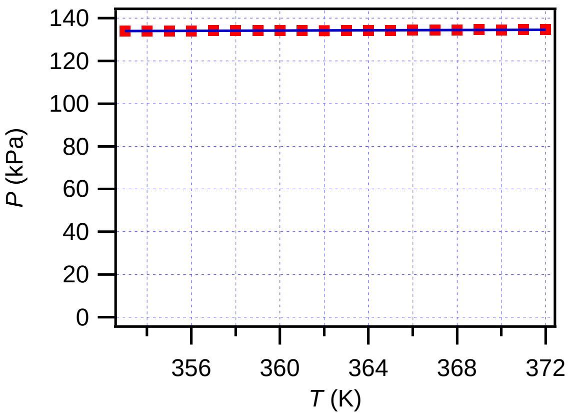
$w = 2.63\ (\pm 0.19)\ x + 1.61\ (\pm 0.5)$

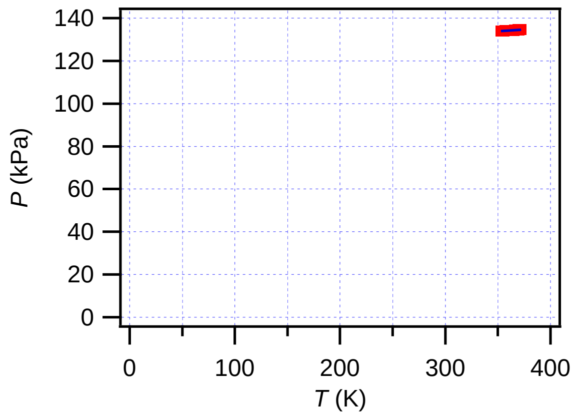
$z = 3.27\ (\pm 0.4)\ x + 2.1\ (\pm 0.9)$

Σωστή γραφική επεξεργασία δεδομένων και Excel

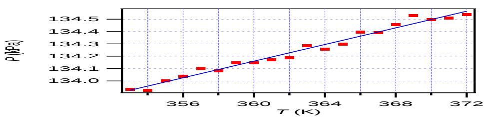
Πολλές φορές τα δεδομένα μας περιγράφονται ικανοποιητικά από μια γραμμική συνάρτηση, δηλ. τα ζεύγη τιμών (x_i, y_i) έχουν τέτοιες τιμές ώστε να υπάρχουν παράμετροι α και β οι οποίες να ικανοποιούν σε μεγάλη προσέγγιση την σχέση $y_i = \alpha\, x + \beta$ για όλα τα δεδομένα με $i = 1, 2, \ldots N$. Όταν σχεδιάζουμε τα δεδομένα σε διάγραμμα, επιλέγουμε το διάστημα τιμών σε κάθε άξονα να είναι το κατάλληλο ώστε τα σημεία να μην περιορίζονται σε μικρό τμήμα, όπως στα επόμενα σχήματα:

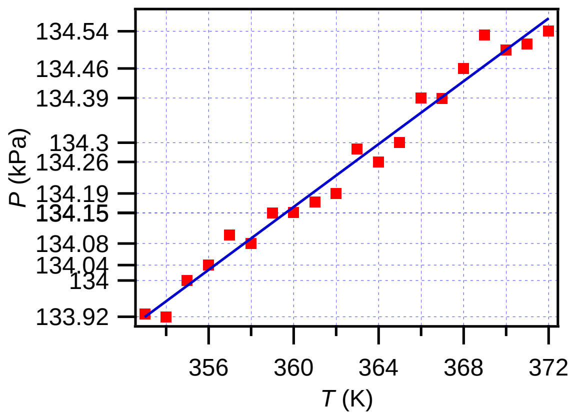






Φροντίζουμε να είναι ευανάγνωστες οι τιμές στους άξονες και σε ισαπέχουσες θέσεις, όχι έτσι:





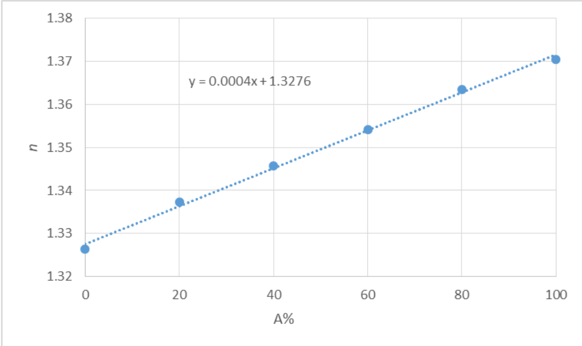
Η καλύτερη ευθεία μπορεί να σχεδιασθεί με οπτική εκτίμηση ή με υπολογισμό (με την μέθοδο ελαχίστων τετραγώνων). Αν την σχεδιάσουμε με οπτική εκτίμηση, προσπαθούμε να ισοκατανείμουμε τα σημεία εκατέρωθεν της ευθείας. Αν κάνουμε υπολογισμό της εξισώσεως που περιγράφει την ευθεία, δεν χαράσσουμε την ευθεία κατ'εκτίμηση, αλλά χρησιμοποιούμε την εξίσωση για να προσδιορίσουμε τις συντεταγμένες δύο σημείων, τα σημειώνουμε στο διάγραμμα και τα ενώνουμε με ευθεία γραμμή.
Η εξίσωση της ευθείας κατά τον οπτικό σχεδιασμό της καλύτερης ευθείας απαιτεί εκτίμηση της κλίσης (α) και του σταθερού όρου (β). Οι δύο παράμετροι έχουν τις κατάλληλες μονάδες ώστε η πράξη $\alpha\, x + \beta$ να δίνει μέγεθος με μονάδες του y. Παρατηρώντας το (σωστό) διάγραμμα διαπιστώνουμε ότι η ευθεία περνά από τα σημεία (355.2 K, 134 kPa) και (370 K, 134.5 kPa).

Άρα $\alpha = \dfrac{(134.5-134.0)\,\text{kPa}}{(370-355.2)\,\text{K}} = \dfrac{0.5\,\text{kPa}}{14.8\,\text{K}} = 0.03378\dfrac{\text{kPa}}{\text{K}}$.

Ο σταθερός όρος υπολογίζεται με την βοήθεια της κλίσης και του ενός εκ των δύο σημείων, π.χ. β = 134.5 kPa – 0.03378 kPa K^{-1} × 370 K = 122.00 kPa.
Δεν μετρούμε τις συντεταγμένες των σημείων σε εκατοστά του μέτρου, ή, αν το κάνουμε, πολλαπλασιάζουμε τις τιμές αυτές με τους συντελεστές κλίμακας, δηλ. πόσα kPa ή K αντιστοιχούν ανά εκατοστό. Αν παραλείψουμε τους συντελεστές, η κλίση έχει τιμή η οποία, αντί να εξαρτάται από τις πειραματικές τιμές, εξαρτάται από την δική μας επιλογή του τρόπου σχεδιάσεως των τιμών στο διάγραμμα.
Με την μέθοδο ελαχίστων τετραγώνων, οι παράμετροι υπολογίζονται ως εξής:
α = 0.03373 kPa K^{-1} και β = 122.02 kPa. Οι τιμές αυτές είναι πολύ κοντά στις τιμές που προσδιορίσαμε γραφικά. Η μέθοδος ελαχίστων τετραγώνων έχει το πλεονέκτημα ότι μας δίνει μια αμερόληπτη (ανεξάρτητη από τον παρατηρητή) τιμή για τις παραμέτρους και παρέχει επίσης τις αβεβαιότητες των παραμέτρων. Στο συγκεκριμένο παράδειγμα, οι αβεβαιότητες είναι σ_α = 0.0012 και σ_β = 0.4.

Κατά την απεικόνιση και επεξεργασία δεδομένων στο Excel, μπορεί να χρησιμοποιηθεί η δυνατότητα αυτόματου υπολογισμού εξισώσεως «γραμμής τάσεως», δηλ. καμπύλη ορισμένου τύπου με παραμέτρους υπολογισμένες με την μέθοδο ελαχίστων τετραγώνων. Τα ακόλουθα δεδομένα δείκτη διαθλάσεως μιγμάτων μεθανόλης και οξικού αιθυλεστέρα μπορούν να παρασταθούν όπως στο σχήμα.

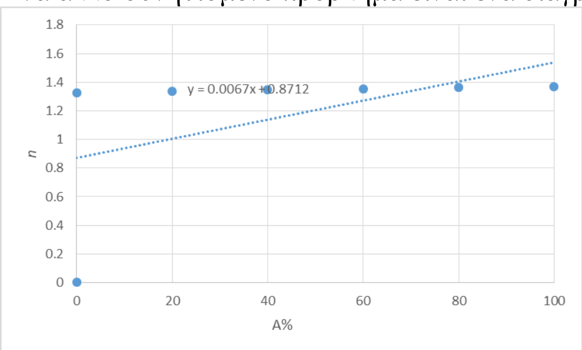
A%	*n*
0	1.3263
20	1.3373
40	1.3457
60	1.3541
80	1.3635
100	1.3705



Η εικονιζόμενη εξίσωση δεν είναι αξιοποιήσιμη διότι δεν εμφανίζει τις παραμέτρους της ευθείας με επαρκή αριθμό σημαντικών ψηφίων και οποιοσδήποτε υπολογισμός βασιζόμενος σε αυτή θα εισαγάγει σοβαρά λάθη. Για να διορθωθεί το πρόβλημα υπάρχουν 2 λύσεις. Η πιο απλή είναι να ζητηθεί η απεικόνιση της εξισώσεως με περισσότερα σημαντικά ψηφία (δεξί κλικ πάνω στην εξίσωση, μορφοποίηση ετικέτας γραμμής τάσης, Αριθμός: κατηγορία: αριθμός με πλήθος δεκαδικών ψηφίων: 6. Τότε η εξίσωση γίνεται $y = 0.000440x + 1.327567$, που είναι αποδεκτό. Η πιο προηγμένη λύση είναι με την χρήση της συναρτήσεως του Excel LINEST. Επιλέγουμε ένα χώρο 10 κελίων, 2 στήλες επί 5 γραμμές και εισάγουμε σε όλα τα πεδία την ίδια συνάρτηση (πατώντας ctrl-shift-Enter αντί για Enter). Γράφουμε =LINEST(B2:B7;A2:A7;TRUE;TRUE), αν τα δεδομένα του οριζόντιου άξονα είναι στις θέσεις A2 ως A7, του κατακόρυφου άξονα στις θέσεις A2 ως B7. Στην επιλεγμένη περιοχή θα εμφανιστούν τα εξής αποτελέσματα. Το πλεονέκτημα αυτής της διαδικασίας είναι ότι τα α και β μπορούν να χρησιμοποιηθούν σε επόμενες πράξεις χωρίς να χρειάζεται να πληκτρολογηθούν μετά από αντιγραφή από το πλαίσιο της εξισώσεως πάνω στο διάγραμμα.

0.00044	1.327567	α	β
1.26E-05	0.000764	σ_α	σ_β
0.996725	0.001055	R^2	σ_y
1217.246	4	F	Βαθμοί ελευθ.
0.001355	4.45E-06	χ^2	adj χ^2

Ένα άλλο συνηθισμένο πρόβλημα είναι ένα διάγραμμα αυτού του τύπου.



Είναι εμφανές ότι η υπολογισμένη και σχεδιασμένη ευθεία δεν είναι ικανοποιητική προσέγγιση των δεδομένων τα οποία φαίνεται να σχηματίζουν μια πολύ καλή ευθεία. Αναζητούμε την αιτία και συνειδητοποιούμε ότι στο διάγραμμα περιλαμβάνονται σημεία που δεν ανήκουν στο σύνολο των μετρήσεων. Δηλ. το διάγραμμα κατασκευάστηκε με επιλογή περισσότερων κελιών από τα απαιτούμενα. Προφανώς και η εξίσωση είναι λανθασμένη.