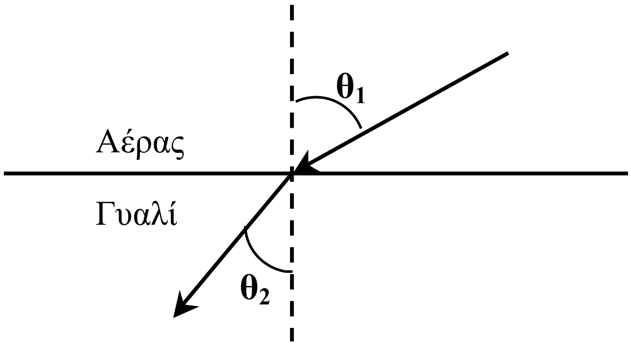
Πρακτική με στοιχεία στατιστικής ανάλυσης

1. Για να υπολογίσουμε μια ποσότητα $q = x^2 y - xy^2$, μετρήσαμε τα μεγέθη x και y και βρήκαμε $x = 3.0 \pm 0.1$ και $y = 2.0 \pm 0.1$. Να βρεθεί η ποσότητα q και η αβεβαιότητά της.
2. Όταν μια ακτίνα φωτός περάσει από τον αέρα σε γυαλί, η γωνία πρόσπτωσης θ_1 και η γωνία διάθλασης θ_2 ορίζονται όπως στο παρακάτω σχήμα. Οι γωνίες σχετίζονται μεταξύ τους σύμφωνα με το νόμο του Snell: $\sin\theta_1 = n\sin\theta_2$, όπου n



είναι ο δείκτης διάθλασης του γυλιού. Επομένως αν μετρήσουμε τις γωνίες θ_1 και θ_2 μπορούμε να βρούμε το δείκτη διάθλασης n, $n = \dfrac{\sin\theta_1}{\sin\theta_2}$. Έστω ότι μετρήσαμε τη γωνία θ_2 για δυο περιπτώσεις της γωνίας πρόσπτωσης θ_1 και η αβεβαιότητα της μέτρησης των δυο γωνιών ήταν 1^{o} ή 0.02rad. Τα αποτελέσματα των μετρήσεων φαίνονται στο παρακάτω πίνακα. Συμπλήρωστε τις θέσεις του πίνακα με τους κατάλληλους υπολογισμούς. Αν ο κατασκευαστής του γυαλιού ισχυρίζεται ότι ο δείκτης διάθλασης του γυαλιού που χρησιμοποιεί είναι n=1.5 θεωρείτε ότι οι μετρήσεις σας είναι συμβατές;

$\theta_1 \left(\pm 1^{o}\right)$	$\theta_2 \left(\pm 1^{o}\right)$	$\sin(\theta_1)$	$\sin(\theta_2)$	η	$\dfrac{\delta\sin\theta_1}{\sin\theta_1}$	$\dfrac{\delta\sin\theta_2}{\sin\theta_2}$	$\dfrac{\delta n}{n}$
10	7						
20	13						
30	20						
50	29						
70	38						

3. Αν είχαμε μετρήσει την ακτίνα μιας σφαίρας σαν $r = 2.0 \pm 0.1m$ τι τιμή θα έπρεπε να αναφέρουμε για τον όγκο της;
4. Τέσσερεις μετρήσεις του μήκους κύματος του φωτός που εκπέμπεται από κάποιο άτομο έχουν γίνει. Τα αποτελέσματα σε nano-meters είναι 503 ± 10, 491 ± 8, 515 ± 20 και 570 ± 40. Να βρεθεί η σταθμισμένη μέση τιμή και η αβεβαιότητά της. Νομίζετε ότι θα πρέπει να συμπεριλάβετε τη τελευταία μέτρηση;
5. Κατά τη διάρκεια δυο ωρών ένας πυρηνικός φυσικός κάνει 12 μετρήσεις της ισχύος μιας ραδιενεργούς πηγής με μεγάλο χρόνο ζωής. Τα αποτελέσματά του σε

μονάδες milli-Curie είναι 12,34,22,14,22,17,24,22,18,14,18,12. Επειδή η πηγή έχει μεγάλο χρόνο ζωής η δραστηριότητα της δε θα αλλάξει αρκετά κατά τη διάρκεια των μετρήσεων. (α) Ποια είναι η μέση τιμή και τυπική απόκλιση των μετρήσεών του; (β) Σύμφωνα με το κριτήριο απόρριψης τιμών του Chauvenet νομίζετε ότι θα πρέπει να απορρίψει τη τιμή 34 ως λάθος μέτρηση; Εξηγήστε τη γνώμη σας. (γ) Αν όντως απορρίψει τη τιμή αυτή ποια είναι η νέα μέση τιμή και τυπική απόκλιση;

6. Αν ο όγκος ενός δείγματος ιδανικού αερίου παραμείνει σταθερός τότε η θερμοκρασία του T είναι γραμμική συνάρτηση της πίεσής του: $T = A + BP$. Στη περίπτωση αυτή η σταθερά A είναι η θερμοκρασία στην οποία η πίεση θα γίνει μηδέν (αν το αέριο δεν μετατραπεί σε υγρό πρώτα). Ονομάζεται απόλυτο μηδέν της θερμοκρασίας και η τιμή της είναι -273.15^o C. Η σταθερά B εξαρτάται από τη φύση του αερίου, τη μάζα του και τον όγκο του. Κάνοντας μια σειρά μετρήσεων για T και P μπορούμε να εξάγουμε τις καλύτερες εκτιμήσεις μας για τις σταθερές A και B.

Μια σειρά μετρήσεων (5 συνολικά) πραγματοποιήθηκαν από κάποιο φοιτητή και δείχνονται στο παρακάτω πίνακα. Ο φοιτητής θεώρησε ότι οι μετρήσεις του για την πίεση P, είχαν αμελητέο σφάλμα ενώ οι αβεβαιότητες ως προς τη θερμοκρασία T όλες είχαν την ίδια αβεβαιότητα (μερικών βαθμών). Υποθέτοντας ότι τα πειραματικά του σημεία προσαρμόζονται σε μια ευθεία γραμμή ελαχίστων τετραγώνων υπολογίζει την σταθερά A και την αβεβαιότητα της και τη σταθερά B και αβεβαιότητά της. Τι ακριβώς έχει υπολογίσει για A και B? Σχεδιάστε τη γραφική παράσταση των σημείων του καθώς και την ευθεία των ελαχίστων τετραγώνων. Εφόσον έχει υπολογίσει τη σταθερά A ξέρει και τη θερμοκρασία του απόλυτου μηδενός. Τι συμπέρασμα θα μπορούσε να εξάγει για την τιμή αυτή ως προς τη θεωρητικά αναμενόμενη;

Μέτρηση	Πίεση (P)	Θερμοκρασία (T)
1	65	-20
2	75	17
3	85	42
4	95	94
5	105	127

7. Ο πληθυσμός διαφόρων ειδών (άνθρωποι, βακτήρια, ραδιανεργοί πυρήνες) μεταβάλονται εκθετικά με το χρόνο. Αν ένας πληθυσμός N στοιχείων ελαττώνεται εκθετικά τότε μπορούμε να γράψουμε $N = N_0 e^{-t/\tau}$, όπου τ είναι ο χρόνος μέσης ζωής του πληθυσμού. Έστω ότι ο πληθυσμός κάποιων βακτηριδίων ελαττώνεται σύμφωνα με τη προηγούμενη σχέση και μετράμε το πληθυσμό για 3 συνεχόμενες ημέρες και πέρνουμε τα αποτελέσματα που φαίνονται στο παρακάτω πίνακα. Με βάση τις μετρήσεις να βρεθεί ο μέσος χρόνος ζωής τ και η αβεβαιότητά του, και να γίνει η γραφική παράσταση σε κατάλληλο χαρτί.

Χρόνος (ημέρες)	Πληθυσμός N_i
0	153,000
1	137,000
2	128,000

Υπόδειξη/Σημείωση: Το πρόβλημα παρόλο που φαίνεται απλό και μπορεί να λυθεί με τις σχέσεις που έχουμε δει για τη μέθοδο των ελαχίστων τετετραγώνων ωστόσο παρουσιάζει μια δυσκολία. Υποθέτουμε ότι όλες οι τιμές N_i έχουν την ίδια αβεβαιότητα σ_{Ni}. Ωστόσο όταν εφαρμόσουμε τη μέθοδο των ελαχίστων τετραγώνων χρησιμοποιούμε $y_i=\ln N_i$ και επομένως το σφάλμα της κάθε τιμής y_i δεν είναι το ίδιο. Επομένως θα πρέπει να χρησιμοποιηθούν οι σχέσεις της μεθόδου ελαχίστων τετραγώνων με διαφορετικές αβεβαιότητες για κάθε μέτρηση. Εξηγήστε γιατί οι αβεβαιότητες δεν είναι ίδιες για όλες τις μετρήσεις και βρείτε την αβεβαιότητα της κάθε μέτρησης.

8. Κάποιος φοιτητής μετρά την ταχύτητα ενός σώματος που κινείται σε αεροδιάδρομο και χρησιμοποιεί μια φωτογραφία πολλαπλών εκθέσεων (αποτυπώνονται στην ίδια φωτογραφία οι θέσεις του σώματος σε διαφορετικές χρονικές στιγμές) για να βρει τις θέσεις s, του σώματος για 5 διαδοχικές ισαπέχουσες χρονικές στιγμές όπως στο πίνακα. (α) Ένας τρόπος για να βρει τη ταχύτητα του σώματος είναι να υπολογίσει $\upsilon = \Delta s/\Delta t$ σε κάθε ένα από τα 4 διαδοχικά χρονικά διαστήματα και μετά να βρει τη μέση τιμή των ταχυτήτων αυτών. Δείξτε ότι με το τρόπο αυτό η ταχύτητα είναι $\upsilon = (s_5 - s_1)/(t_5 - t_1)$, που σημαίνει ότι οι ενδιάμεσες τρεις μετρήσεις αγνοούνται τελείως. Αποδείξτε το χωρίς να χρησιμοποιήσετε αριθμητικές τιμές και κατόπιν βρείτε την αριθμητική απάντηση. (β) Μια καλύτερη μέθοδος είναι να εφαρμόσει τη μέθοδο των ελαχίστων τετραγώνων και να προσαρμόσει τα δεδομένα του στην εξίσωση $s = s_0 - \upsilon t$ χρησιμοποιώντας όλα τα δεδομένα. Ακολουθώντας τη μέθοδο αυτή βρείτε τη καλύτερη τιμή της ταχύτητας υ και συγκρίνετε τα αποτελέσματά σας με αυτό της (α) μεθόδου (Θεωρήστε ότι η αβεβαιότητα στο χρόνο είναι αμελητέα και ότι οι αβεβαιότητες της μέτρησης της θέσης είναι όλες ίδιες)

Χρόνος, t(sec)	-4	-2	0	2	4
Θέση, s(cm)	13	25	34	42	56

9. Αν έχετε μια πολύ καλή εκτίμηση της αβεβαιότητας δy των μετρήσεών σας y_i τότε μπορείτε να συγκρίνετε την εκτίμηση αυτή με τη τυπική απόκλιση σ_y που προκύπτει από τη μέθοδο των ελαχίστων τετραγώνων ($\sigma_y = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - a - bx_i)^2}$) και να εκτιμήσετε κατά πόσο τα δεδομένα σας επαληθεύουν/συμφωνούν με την αναμενόμενη γραμμική εξάρτηση y = A + Bx. Η ποσότητα σ_y είναι περίπου η μέση απόσταση κατά τη οποία η μέτρηση (x_i,y_i) δεν κατορθώνει να πέσει στη γραμμή των ελαχίστων τετραγώνων. Αν η εκτίμηση της σ_y είναι περίπου ίδια με την εκτιμούμενη αβεβαιότητα δ_y, τα δεδομένα είναι συμβατά με την αναμενόμενη γραμμική σχέση. Αν όμως σ_y, είναι πολύ μεγαλύτερη από δ_y, υπάρχει σημαντικός λόγος να πιστεύουμε ότι η γραμμική εξάρτηση δεν ισχύει. Για παράδειγμα έστω ότι ένας φοιτητής πέρνει μετρήσεις της θέσης ενός σώματος σε ίσα διαδοχικά χρονικά διαστήματα όπως στο προηγούμενο πρόβλημα. Υποθέτοντας ότι το σώμα κινείται με σταθερή ταχύτητα προσαρμόζει τα δεδομένα του με τη γραμμική εξίσωση $s = s_0 + \upsilon t$. (α) Χρησιμοποιήστε τη μέθοδο των ελαχίστων τετραγώνων για να βρείτε τα s_0, υ και σ_s των μετρήσεων s. (β) Υποθέστε ότι η αβεβαιότητα στη μέτρηση της θέσης του σώματος ήταν δs=1cm. Συγκρίνοντας τη τιμή δ_s με τη τιμή σ_s, δώστε κάποιο επιχείρημα για το αν τα δεδομένα συνάδουν με την γραμμική εξάρτηση και την υπόθεση ότι η ταχύτητα είναι σταθερή. Σχεδιάστε το γράφημα της θέσης-χρόνου για τις μετρήσεις και τα αντίστοιχα σφάλματα των μετρήσεων καθώς και την ευθεία προσαρμογής για να επιβεβαιώσετε το συμπέρασμά σας. (γ) Υποθέστε τώρα ότι ο φοιτητής είναι σίγουρος ότι η αβεβαιότητα των μετρήσεών του είναι όχι περισσότερο από 0.1cm. Στη περίπτωση αυτή νομίζετε ότι οι μετρήσεις του είναι συμβατές με την υπόθεση της κίνησης με σταθερή ταχύτητα; Αφού εξετάσετε το γράφημά σας μπορείτε να δώσετε κάποια εξήγηση για το τι μπορεί να συμβαίνει;

Χρόνος, t(sec)	-3	-1	1	3
Θέση, s(cm)	4.0	7.5	10.3	12.0

10. Στην άσκηση αυτή υπάρχουν πολλά στοιχεία από τη θεωρία και θα ήταν καλό να τα διαβάσετε προσεκτικά. Το τι σας ζητά να κάνετε είναι μετά τη θεωρητική εισαγωγή που σας δίνεται.

Συζητήσαμε στη τάξη τη περίπτωση που δυο μεγέθη x και y που εξετάζουμε μπορεί να έχουν κάποιο βαθμό συσχέτισης η γενική μορφή της αβεβαιότητας ενός μεγέθους Q που εξαρτάται από τα x και y πέρνει τη μορφή:

$$\sigma_Q^2 = \left(\frac{\partial Q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial Q}{\partial y}\right)^2 \sigma_y^2 + 2\left(\frac{\partial Q}{\partial x}\right)\left(\frac{\partial Q}{\partial y}\right)\sigma_{xy} \qquad \textbf{(1)}$$

όπου ο όρος σ_{xy} ονομάζεται ***συναλλοίωτος*** και δίνεται από τη σχέση:

$$\operatorname{cov}(x,y) = \sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right) \qquad \textbf{(2)}$$

Στη περίπτωση που τα ζεύγη των μετρήσεών μας συνδέονται με κάποια γραμμική εξάρτηση της μορφής $y = A + Bx$ (όπως στη περίπτωση της ευθείας των ελαχίστων τετραγώνων), τότε η συσχέτιση των μεγεθών x και y εκφράζεται από το ***παράγοντα γραμμικής συσχέτισης*** ή απλά ***παράγοντας συσχέτισης*** και δίνεται από τη σχέση:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2 \sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2}} \qquad \textbf{(3)}$$

όπου τα σ_x, σ_y και σ_{xy} είναι η τυπική απόκλιση του x, y και η συναλλοίωτος όρος των x και y. Ο παράγοντας συσχέτισης r, έχει τιμές μεταξύ -1 και 1, με -1 πλήρως αντισχετιζόμενες μεταβλητές x και y, +1 πλήρως σχετιζόμενες ενώ για r=0 οι μεταβλητές είναι τελείως ανεξάρτητες μεταξύ τους.

Για παράδειγμα θεωρήστε ότι όλα τα ζεύγη τιμών (x_i,y_i) βρίσκονται πάνω σε μια ευθεία y=A+Bx. Στη περίπτωση αυτή για όλα τα ζεύγη i θα έχουμε $y_i=A+Bx_i$ και επομένως $\overline{y} = A + B\overline{x}$. Αφαιρώντας τις δυο εξισώσεις μεταξύ τους για κάθε ζεύγος i, θα έχουμε $y_i - \overline{y} = B\left(x_i - \overline{x}\right)$. Απλή αντικατάσταση στην εξίσωση του παράγοντα συσχέτισης παραπάνω θα δώσει:

$$r = \frac{B\sum_{i=1}^{N}(x_i-\bar{x})^2}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2 B^2 \sum_{i=1}^{N}(x_i-\bar{x})^2}} = \frac{B}{|B|} = \pm 1 \qquad \textbf{(4)}$$

Δηλαδή όλα τα σημεία θα βρίσκονται πάνω στην ευθεία και $r = \pm 1$ με το πρόσημό του να καθορίζεται από τη κλίση της ευθείας.

Το τελευταίο που θα πρέπει να σημειώσουμε σχετικά με σχετιζόμενες ή μη μεταβλητές είναι η περίπτωση που οι μετρήσεις που έχουμε είναι πεπερασμένες. Ας υποθέσουμε ότι οι μεταβλητές x και y είναι ***μή σχετιζόμενες***. Ανεξάρτητα από τη τιμή y_i κάθε x_i θα είναι το ίδιο πιθανό να έχει τιμή μεγαλύτερη από $\bar{x}$ ή μικρότερη από $\bar{x}$. Επομένως ο αριθμητής της εξίσωσης (3) είναι το ίδιο πιθανό να είναι θετικός ή αρνητικός ενώ ο παρονομαστής είναι πάντοτε θετικός. Στη περίπτωση που ο αριθμός των μετρήσεων γίνεται πολύ μεγάλος $N \to \infty$, ο παράγοντας συσχέτισης θα γίνει r=0. Για πεπερασμένες μετρήσεις r θα είναι μικρό αλλά όχι μηδέν. Το πρόβλημα που έχουμε είναι πως ξέρουμε ότι ο παράγοντας r που υπολογίζουμε από N πεπερασμένες μετρήσεις και είναι διαφορετικός από μηδέν δεν αντιστοιχεί σε μη σχετιζόμενα μεγέθη που εξαιτίας της πεπερασμένης στατιστικής μας έδωσαν τιμή διαφορετική από 0. Δηλαδή αν μετρήσουμε r=0.6 (τα μεγέθη είναι αρκετά σχετιζόμενα μεταξύ τους) ποια είναι η πιθανότητα τα μεγέθη αυτά να ήταν όντως ανεξάρτητα μεταξύ τους αλλά επειδή ο αριθμός των μετρήσεων ήταν πολύ μικρός έτυχε να μετρήσουμε τη τιμή r=0.6. Για το λόγο αυτό ορίζουμε τη πιθανότητα $P_N\left(|r| \geq r_0\right)$ που εκφράζει τη πιθανότητα δυο μη σχετιζόμενες μεταβλητές x και y θα δώσουν τιμή παράγοντα συσχέτισης r μεγαλύτερη από την συγκεκριμένη τιμή r_0 την οποία λάβαμε με τις N πεπερασμένες τιμές των μεγεθών που έχουμε.

Ο παρακάτω πίνακας δίνει τη πιθανότητα $P_N\left(|r| \geq r_0\right)$ για N μετρήσεις δυο μη σχετιζόμενων μεταβλητών x και y να οδηγήσουν σε παράγοντα συσχέτισης r με τιμή $|r| \geq r_0$ Οι τιμές του πίνακα είναι % πιθανότητες ενώ κενές θέσεις αντιστοιχούν σε τιμές πιθανότητας μικρότερες από 0.05%.

N	r_0										
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
3	100	94	87	81	74	67	59	51	41	29	0
6	100	85	70	56	43	31	21	12	6	1	0
10	100	78	58	40	25	14	7	2	0.5		0
20	100	67	40	20	8	2	0.5	0.1			0
50	100	49	16	3	0.4						0

Οι αριθμοί σε κάθε στήλη δείχνουν την επι τοις εκατό πιθανότητα N μετρήσεις 2 μη σχετιζόμενων μεταβλητών x και y να δώσουν τιμή r τουλάχιστον τόσο μεγάλη όσο η τιμή r_0 που αναγράφεται στην αρχή κάθε στήλης. Για παράδειγμα η πιθανότητα για 10 μετρήσεις των ανεξάρτητων μεταβλητών x και y να δώσουν τιμή $|r| \geq 0.8$ είναι 0.5%. Συνήθως θεωρούμε ότι πιθανότητα της τάξης 0.5% ή μικρότερη σημαίνει ότι οι δυο μεταβλητές δεν είναι ανεξάρτητες και υπάρχει συσχέτιση η οποία είναι σημαντική.

Μετά την εισαγωγή αυτή, υπολογίστε το παράγοντα συσχέτισης r για τα ακόλουθα 6 ζεύγη μετρήσεων:

x	1	2	3	5	6	7
y	5	6	6	8	8	9

Πιστεύετε ότι άν υποθέτατε ότι τα μεγέθη ήταν ανεξάρτητα μεταξύ τους η πιθανότητα να πάρετε τη συγκεκριμένη τιμή που υπολογίσατε αποδεικνύει ότι τα μεγέθη είναι όντως ανεξάρτητα μεταξύ τους;