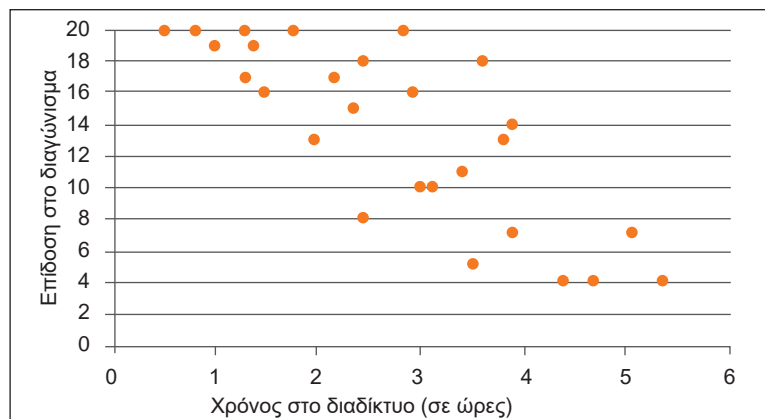
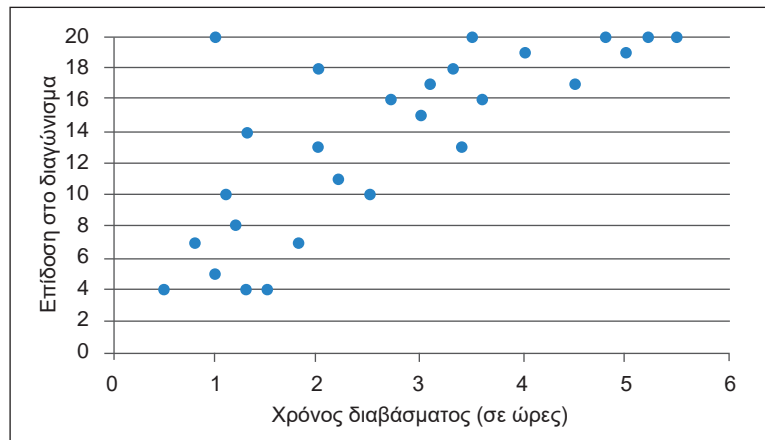


ΕΝΟΤΗΤΑ 2.7 ΓΡΑΜΜΙΚΗ ΣΥΣΧΕΤΙΣΗ ΠΟΣΟΤΙΚΩΝ ΜΕΤΑΒΛΗΤΩΝ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΔΙΑΣΠΟΡΑΣ

Διερεύνηση

Τα επόμενα διαγράμματα παρουσιάζουν την επίδοση 27 μαθητών ενός τμήματος της Β΄ τάξης του Γενικού Λυκείου στην Άλγεβρα σε σχέση με τον χρόνο που αφιέρωσαν τις προηγούμενες δύο μέρες για διάβασμα και ενασχόληση στο διαδίκτυο.



- α) Ποιες είναι οι μεταβλητές του πρώτου και δεύτερου διαγράμματος; Ποια από τις δύο μπορεί να θεωρηθεί ως ανεξάρτητη και ποια ως εξαρτημένη από την άλλη;
- β) Να συγκρίνετε τη σχέση των μεταβλητών «επίδοση στο διαγώνισμα» και «χρόνος διαβάσματος» με τη σχέση των μεταβλητών «επίδοση στο διαγώνισμα» και «χρόνος στο διαδίκτυο».
- γ) Να χαράξετε «με το μάτι» μια ευθεία που να προσαρμόζεται καλύτερα στα δεδομένα.
- δ) Για ποιον λόγο θα χρησιμοποιούσατε μια ευθεία του γ) ερωτήματος;

Βασικές Μαθηματικές Έννοιες – Ιδέες – Διεργασίες

Έχουμε δει μέχρι τώρα ότι η σχέση δύο ποιοτικών μεταβλητών περιγράφεται με τη βοήθεια των πινάκων συνάφειας και των κατάλληλων ραβδογραμμάτων, ενώ υπολογίζοντας τα μέτρα θέσης και μεταβλητότητας μιας ποσοτικής μεταβλητής στις κατηγορίες μιας ποιοτικής μπορούμε να συγκρίνουμε ομάδες δεδομένων. Τι συμβαίνει εάν έχουμε δύο ποσοτικές μεταβλητές; Πώς μπορούμε να περιγράψουμε τη σχέση δύο ποσοτικών μεταβλητών; Για παράδειγμα:

- Η ηλικία και το βάρος ενός παιδιού έχουν κάποια θετική συσχέτιση μεταξύ τους, με την έννοια ότι όσο πιο μεγάλο είναι το παιδί τόσο μεγαλύτερο βάρος θα έχει.
- Η διάρκεια ζωής των ζώντων οργανισμών σε μια περιοχή και ένας δείκτης μόλυνσης της περιοχής έχουν αρνητική συσχέτιση μεταξύ τους, με την έννοια ότι όσο πιο μεγάλος είναι ο δείκτης μόλυνσης της περιοχής τόσο μικρότερη είναι η διάρκεια ζωής των οργανισμών που ζουν στην περιοχή.
- Το ύψος των αποδοχών των υπαλλήλων μιας εταιρείας δε συσχετίζεται με το βάρος τους.

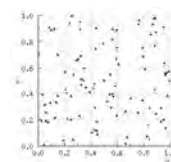
Στην ενότητα αυτή θα εξετάσουμε μόνο την περίπτωση της γραμμικής συσχέτισης μεταξύ δύο ποσοτικών μεταβλητών. Για παράδειγμα, καταγράφουμε το ύψος (X σε cm) και το βάρος (Y σε kg) 18 μαθητών της Β΄ Λυκείου. Αυτό που μας ενδιαφέρει είναι τι συμβαίνει με το βάρος των παιδιών όταν αλλάζει το ύψος τους ή τι συμβαίνει με το ύψος των παιδιών όταν αλλάζει το βάρος τους. Αν μας ενδιαφέρει η πρώτη περίπτωση, θεωρούμε ως ανεξάρτητη μεταβλητή το ύψος των μαθητών και ως εξαρτημένη (από το ύψος) μεταβλητή το βάρος των μαθητών. Στη δεύτερη περίπτωση συμβαίνει το αντίθετο.

Μαθητής	Ύψος (X)	Βάρος (Y)	Μαθητής	Ύψος (X)	Βάρος (Y)
A	170	58	K	178	79
B	172	60	Λ	179	76
Γ	173	67	M	180	79
Δ	175	72	N	180	80
E	176	65	Ξ	180	83
Z	177	81	O	180	85
H	178	73	Π	182	89
Θ	178	74	P	187	90
I	178	78	Σ	191	92

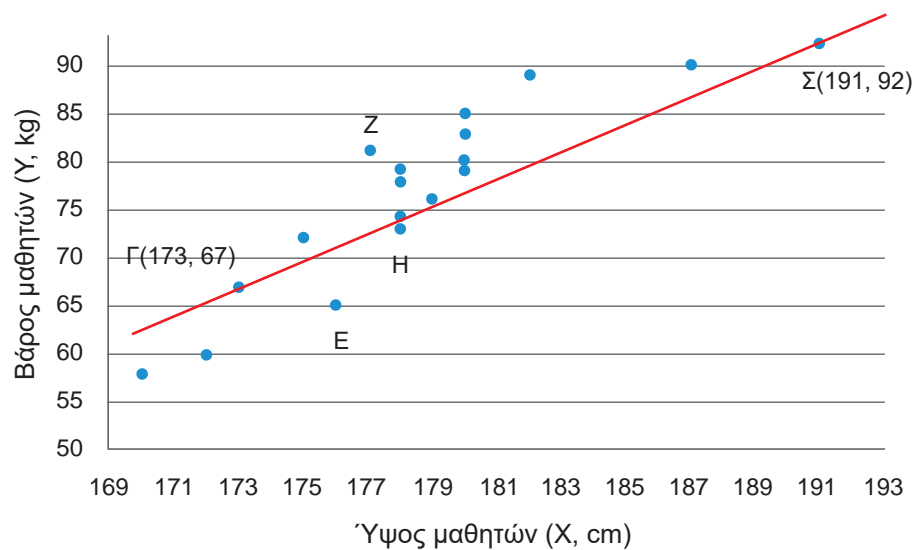
Πίνακας 1: Ύψη (X) και Βάρη (Y) 18 μαθητών της Β΄ Λυκείου

Για κάθε άτομο έχουμε δύο μετρήσεις: το ύψος και το βάρος του. Το δείγμα μας δηλαδή αποτελείται από τα ζεύγη τιμών (x_i, y_i) , $i = 1, 2, \dots, 18$ των συνεχών μετα-

Διάγραμμα Διασποράς



βλητών X και Y . Η γραφική παράσταση των σημείων σε ένα ορθογώνιο σύστημα αξόνων ονομάζεται **διάγραμμα διασποράς** (scatterplot) των μεταβλητών X και Y . Στο διάγραμμα παρατηρούμε μια διασπορά των σημείων που αντιστοιχούν στους μαθητές που εξετάζουμε. Από το διάγραμμα διασποράς μπορούμε να εντοπίσουμε την ύπαρξη συσχέτισης που ενδεχομένως να υπάρχει μεταξύ των μεταβλητών που εξετάζουμε. Η γενική εικόνα που παίρνουμε από το διάγραμμα διασποράς είναι ότι οι ψηλότεροι μαθητές είναι συνήθως και βαρύτεροι. Για παράδειγμα, ο μαθητής Z είναι ψηλότερος και βαρύτερος από τον μαθητή E , ο H είναι ψηλότερος και βαρύτερος από τον E , αλλά υπάρχουν και εξαιρέσεις, όπως ο H είναι ψηλότερος από τον Z αλλά ο Z είναι βαρύτερος από τον H . Το άπλωμα των σημείων στο διάγραμμα θα μπορούσαμε να πούμε ότι γίνεται κοντά σε μια νοητή ευθεία γραμμή, οπότε η σχέση θεωρείται γραμμική.



Σχήμα 1: Διάγραμμα Διασποράς Ύψους και Βάρους 18 μαθητών της Β' λυκείου και ευθεία προσαρμοσμένη «με το μάτι» για τα δεδομένα του πίνακα 1

**Συντελεστής
γραμμικής συσχέτισης
των ποσοτικών
μεταβλητών X και Y**

Ένα μέτρο που μας δίνει το μέγεθος της γραμμικής σχέσης των δύο ποσοτικών μεταβλητών X και Y είναι ο συντελεστής γραμμικής συσχέτισης των X και Y (linear correlation coefficient ή Pearson's correlation coefficient). Ο συντελεστής γραμμικής συσχέτισης δύο ποσοτικών μεταβλητών X και Y ορίζεται με βάση ένα δείγμα n ζευγών παρατηρήσεων (x_i, y_i) , $i = 1, 2, \dots, n$, συμβολίζεται με $r(X, Y)$ ή απλά r και δίνεται από τον τύπο:

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{s_X s_Y}}$$

(1)

όπου \bar{x}, s_x η δειγματική μέση τιμή και δειγματική τυπική απόκλιση της μεταβλητής X και \bar{y}, s_y η δειγματική μέση τιμή και δειγματική τυπική απόκλιση της μεταβλητής Y. Ωστόσο, σε πραγματικές μελέτες με πολλά δεδομένα, η χρήση λογιστικού φύλλου ή κάποιου στατιστικού λογισμικού κρίνεται απαραίτητη.

Ο συντελεστής γραμμικής συσχέτισης δύο ποσοτικών μεταβλητών X και Y είναι καθαρός αριθμός, δηλαδή δεν εκφράζεται σε συγκεκριμένες μονάδες μέτρησης, επομένως είναι ανεξάρτητος των χρησιμοποιούμενων μονάδων μέτρησης των μεταβλητών X και Y. Ισχύει πάντοτε:

$$-1 \leq r \leq 1$$

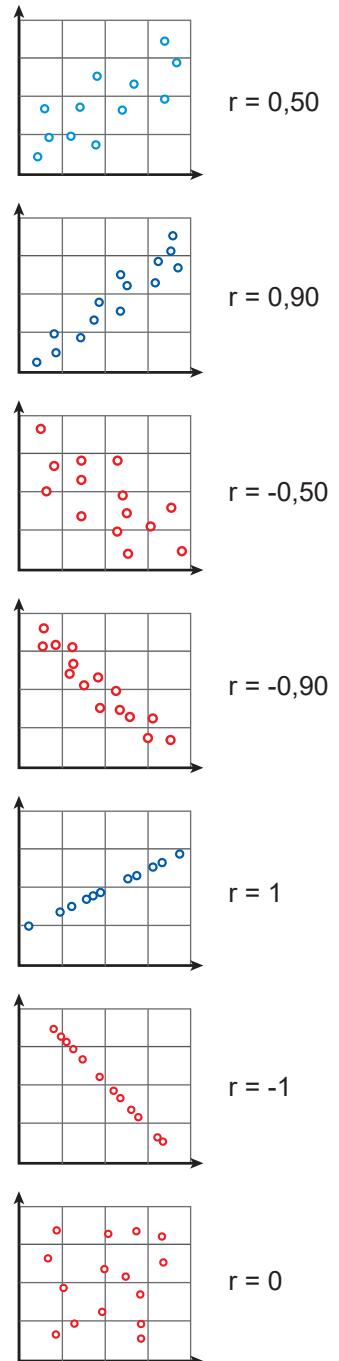
Πιο συγκεκριμένα όταν:

- $0 < r < 1$, τότε οι X και Y είναι **θετικά γραμμικά συσχετισμένες**. Αυτό σημαίνει ότι, όταν οι τιμές της μεταβλητής X αυξάνονται, οι τιμές της Y τείνουν να αυξάνονται. Όσο ο συντελεστής πλησιάζει το 1 τόσο πιο συγκεντρωμένες γύρω από μια νοητή ευθεία είναι οι παρατηρήσεις. Επομένως, τόσο πιο ισχυρή είναι η θετική γραμμική συσχέτιση.
- $-1 < r < 0$, τότε οι X και Y είναι **αρνητικά γραμμικά συσχετισμένες**. Αυτό σημαίνει ότι, όταν οι τιμές της μεταβλητής X αυξάνονται, οι τιμές της Y τείνουν να μειώνονται. Όσο ο συντελεστής πλησιάζει το -1 τόσο πιο συγκεντρωμένες γύρω από μια νοητή ευθεία είναι οι παρατηρήσεις. Επομένως, τόσο πιο ισχυρή είναι η αρνητική γραμμική συσχέτιση.
- $r = 1$, τότε οι X και Y είναι **τέλεια θετικά γραμμικά συσχετισμένες** και όλα τα σημεία βρίσκονται πάνω σε μια ευθεία με θετική κλίση.
- $r = -1$, τότε οι X και Y είναι **τέλεια αρνητικά γραμμικά συσχετισμένες** και όλα τα σημεία βρίσκονται πάνω σε μια ευθεία με αρνητική κλίση.
- $r = 0$, τότε οι X και Y είναι **γραμμικά ασυσχέτιστες, χωρίς αυτό να σημαίνει ότι δεν μπορεί να έχουν κάποια άλλη σχέση, μη γραμμική**.

Από τα δεδομένα του παραδείγματος με τη χρήση αριθμομηχανής ή λογιστικού φύλλου έχουμε:

$$\sum_{i=1}^{18} x_i y_i = 247324, \bar{x} = 178,56, \bar{y} = 76,72, s_x = 4,809, s_y = 9,473,$$

οπότε ο συντελεστής γραμμικής συσχέτισης του ύψους (X) και του βάρους (Y) 18 μαθητών της Β Λυκείου είναι $r = 0,90$, που σημαίνει ότι το ύψος και το βάρος των μαθητών είναι ισχυρά θετικά γραμμικά συσχετισμένες μεταβλητές.



Γραμμική συσχέτιση ποσοτικών μεταβλητών και αιτιότητα

Στο σημείο αυτό θα υπενθυμίσουμε ότι η ισχυρή συσχέτιση μεταξύ δύο μεταβλητών δε συνεπάγεται υποχρεωτικά μια αιτιολογική σχέση ανάμεσα στις δύο μεταβλητές. Υπάρχει περίπτωση η αλλαγή της μεταβλητής X να προκαλεί άμεση αλλαγή της μεταβλητής Y . Αλλά πολύ συχνά οι αλλαγές των δύο μεταβλητών X και Y οφείλονται σε κάποιες άλλες μεταβλητές ή σε κάποιους αστάθμητους παράγοντες. Για παράδειγμα, το μέγεθος παπουτσιού που φορά ένας μαθητής και το μέγεθος των αρχείων που αποθηκεύονται στον υπολογιστή του, ενώ έχουν εξαιρετικά μεγάλη γραμμική συσχέτιση, προφανώς δεν έχουν αιτιώδη σχέση, αλλά επηρεάζονται από έναν τρίτο παράγοντα, την ηλικία (συγχυτικός παράγοντας).

Χάραξη ευθείας στο διάγραμμα διασποράς «με το μάτι»

Τα σημεία (x_i, y_i) , $i=1,2,\dots,18$ είναι συγκεντρωμένα γύρω από μια ευθεία, όπως φαίνεται στο σχήμα 1. Στην πραγματικότητα, σχεδόν ποτέ δεν υπάρχει ευθεία που να περνάει απ' όλα τα σημεία των δεδομένων. Για παράδειγμα, στο πρόβλημά μας με το ύψος και το βάρος των μαθητών, η αλήθεια είναι ότι η μεταβλητή Y (το βάρος των μαθητών) μπορεί να επηρεάζεται και από άλλους παράγοντες εκτός της X (το ύψος των μαθητών), όπως, π.χ., τις διατροφικές συνήθειες. Γι' αυτό μαθητές με το ίδιο ύψος μπορεί να έχουν διαφορετικό βάρος.

Ψάχνουμε, λοιπόν, για εκείνη την ευθεία που περνάει κοντά από τα περισσότερα σημεία των δεδομένων. Όπως γνωρίζουμε, η εξίσωση μιας ευθείας δίνεται από τη σχέση:

$$y = a + \beta x, \quad (2)$$

όπου a και β είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή, όπως λέμε, «να εκτιμήσουμε», έτσι ώστε η ευθεία που θα προκύψει από τη (2) να περιγράφει την αναμενόμενη τιμή y της μεταβλητής Y , όταν η τιμή της μεταβλητής X είναι ίση με x . Η παράμετρος a μας δίνει το σημείο $(0, a)$ όπου η ευθεία αυτή τέμνει τον άξονα $y'y$, ενώ η παράμετρος β παριστάνει τον συντελεστή διεύθυνσης της ευθείας. Ο πιο εύκολος τρόπος χάραξης της ευθείας είναι αυτός που γίνεται «με το μάτι». Μια τέτοια ευθεία έχουμε φέρει στο διάγραμμα διασποράς του σχήματος 1. Για να βρούμε τα a και β επιλέγουμε δύο σημεία, έστω τα $\Gamma(173,67)$ και $\Sigma(191,92)$ πάνω στην ευθεία που φέραμε «με το μάτι» και καταλήγουμε στην εξίσωση:

$$y = -173,28 + 1,39x \quad (3)$$

Επομένως, η ευθεία που κατά τη γνώμη μας προσαρμόζεται καλύτερα στα σημεία του διαγράμματος διασποράς διέρχεται από το σημείο $(0, -173,28)$ και έχει συντελεστή διεύθυνσης 1,39.

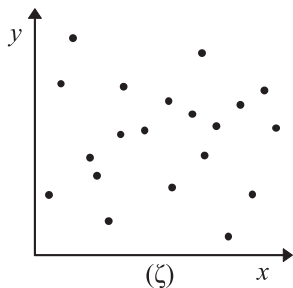
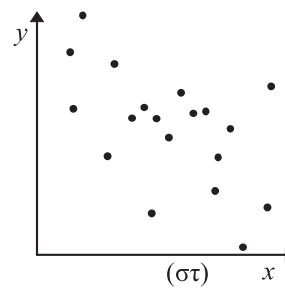
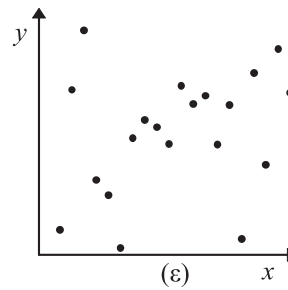
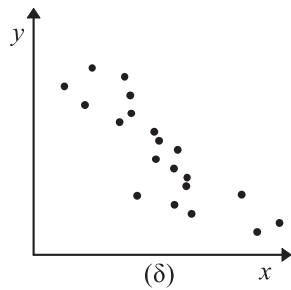
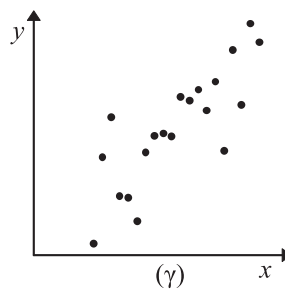
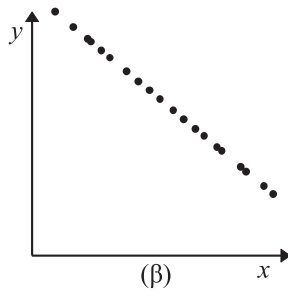
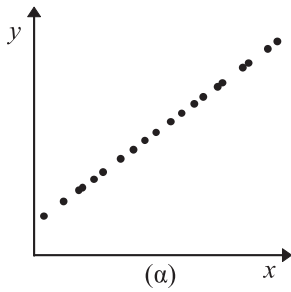
Την εξίσωση της ευθείας (3) μπορούμε να τη χρησιμοποιήσουμε για να εκτιμήσουμε μια τιμή της μεταβλητής Y , από κάποια τιμή της μεταβλητής X που ανήκει στο εύρος τιμών της τελευταίας. Αυτό συμβαίνει, διότι –για τιμές μικρότερες ή μεγαλύτερες από αυτές που παρατηρήθηκαν– δεν είμαστε σίγουροι ότι η γραμμική σχέση παραμένει. Για παράδειγμα: ένας μαθητής με ύψος 1,81 εκτιμάται ότι θα έχει βάρος ίσο με $y = -173,28 + 1,39 \cdot 181 = 78,1$ κιλά.

Εφαρμογές

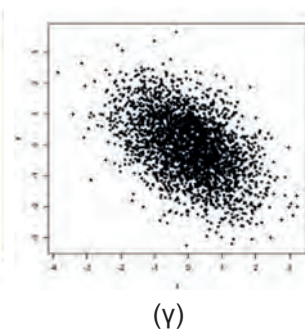
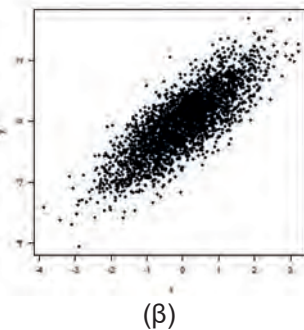
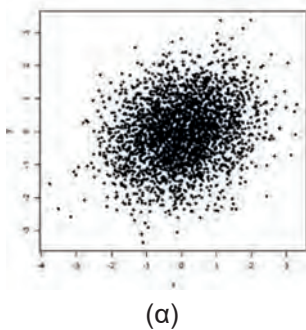
Εφαρμογή 1

A) Για δύο ποσοτικές μεταβλητές X και Y να αντιστοιχίσετε τους συντελεστές γραμμικής συσχέτισης

$r_1 = -1, r_2 = -0,8, r_3 = -0,2, r_4 = 0, r_5 = 0,2, r_6 = 0,8$ και $r_7 = 1$, με τα αντίστοιχα διαγράμματα διασποράς.



B) Για δύο ποσοτικές μεταβλητές X και Y να αντιστοιχίσετε τους συντελεστές γραμμικής συσχέτισης $r_1=0,8, r_2=0,2$ και $r_3= -0,5$, με τα αντίστοιχα διαγράμματα διασποράς.



Λύση

- A)** Σύμφωνα με τις τιμές που παίρνει ένας συντελεστής γραμμικής συσχέτισης μεταξύ δύο ποσοτικών μεταβλητών προκύπτει ότι η τιμή του $r_1 = -1$ αντιστοιχεί στο διάγραμμα (β), η τιμή του $r_2 = -0,8$ στο διάγραμμα (δ), η τιμή του $r_3 = -0,2$ στο διάγραμμα (στ), η τιμή του $r_4 = 0$ στο διάγραμμα (ζ), η τιμή του $r_5 = 0,2$ στο διάγραμμα (ε), η τιμή του $r_6 = 0,8$ στο διάγραμμα (γ) και η τιμή του $r_7 = 1$ στο διάγραμμα (α).
- B)** Σε καθένα από αυτά τα διαγράμματα διασποράς βλέπουμε ένα σμήνος σημείων σε αντίθεση με τα διαγράμματα διασποράς του A που είχαμε λίγα σημεία. Το διάγραμμα διασποράς (α) παρουσιάζει ασθενή θετική γραμμική συσχέτιση με συντελεστή γραμμικής συσχέτισης $r_2=0,2$. Το διάγραμμα διασποράς (β) παρουσιάζει ισχυρή θετική γραμμική συσχέτιση με συντελεστή γραμμικής συσχέτισης $r_1=0,8$. Τέλος, το διάγραμμα διασποράς (γ) παρουσιάζει αρνητική γραμμική συσχέτιση με συντελεστή γραμμικής συσχέτισης $r_3 = -0,5$.

Εφαρμογή 2



Στον Πίνακα 2 δίνονται τα προσδόκιμα ζωής γυναικών (Π.Ζ.Γ.) και ανδρών (Π.Ζ.Α.) του έτους 2005 για 30 τυχαία επιλεγμένες χώρες.

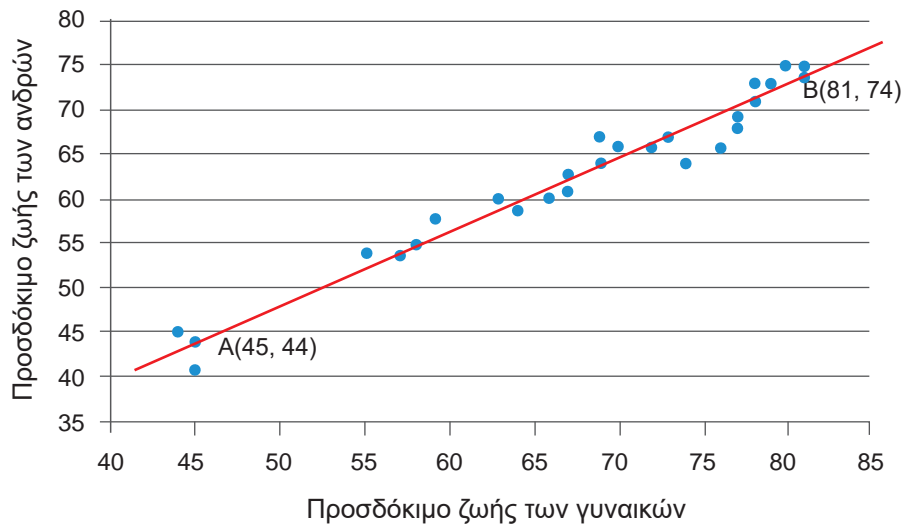
Χώρα	Π.Ζ.Γ.	Π.Ζ.Α.	Χώρα	Π.Ζ.Γ.	Π.Ζ.Α.
Αίγυπτος	63	60	Λευκορωσία	76	66
Αυστρία	79	73	Λιθουανία	77	68
Αφγανιστάν	44	45	Μαλαισία	72	66
Βέλγιο	79	73	Μποτσουάνα	66	60
Βολιβία	64	59	Νησιά Μπαρμπάντος	78	73
Δομινικανή Δημοκρατία	70	66	Νιγηρία	57	54
Ελ Σαλβαδόρ	69	64	Νικαράγουα	67	61
Ελλάδα	80	75	Ολλανδία	81	75
Ζάμπια	45	44	Περου	67	63
Ινδία	59	58	Ρωσία	74	64
Ισημερινός	73	67	Σενεγάλη	58	55
Καμερούν	58	55	Σομαλία	55	54
Καναδάς	81	74	Τανζανία	45	41
Κίνα	69	67	Τσεχία	77	69
Κουβέιτ	78	73	Χιλή	78	71

Πίνακας 2: Προσδόκιμα ζωής των γυναικών (Π.Ζ.Γ.) και ανδρών (Π.Ζ.Α.) για 30 χώρες

- α) Να κατασκευάσετε το διάγραμμα διασποράς μεταξύ των προσδόκιμων ζωής γυναικών και ανδρών και να περιγράψετε το είδος της σχέσης που φαίνεται να έχουν οι δύο μεταβλητές.
- β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των προσδόκιμων ζωής γυναικών και ανδρών.
- γ) Να σχεδιάσετε «με το μάτι» την ευθεία που φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα.
- δ) Αν για την Αλβανία, το προσδόκιμο ζωής των γυναικών είναι τα 75 έτη, μπορείτε να εκτιμήσετε το προσδόκιμο ζωής των ανδρών;

Λύση

- α) Θεωρούμε ως ανεξάρτητη μεταβλητή (X) το προσδόκιμο ζωής των γυναικών και ως εξαρτημένη μεταβλητή (Y) το προσδόκιμο ζωής των ανδρών για 30 χώρες. Βέβαια, στην περίπτωση μας είναι τέτοια η φύση των μεταβλητών που θα μπορούσε να είναι και ανάποδα.



Σχήμα 2: Διάγραμμα Διασποράς Π.Ζ.Γ και Π.Ζ.Α. για 30 χώρες και ευθεία προσαρμοσμένη «με το μάτι» για τα δεδομένα του πίνακα 2

Από το διάγραμμα διασποράς φαίνεται ότι τα σημεία του διαγράμματος είναι κοντά σε μια νοητή ευθεία. Παρατηρούμε ότι, καθώς αυξάνεται το προσδόκιμο ζωής των γυναικών, φαίνεται να αυξάνεται και το προσδόκιμο ζωής των ανδρών.

- β) Ο συντελεστής γραμμικής συσχέτισης υπολογίζεται από τη σχέση (1). Επομένως, υπολογίζουμε πρώτα τη μέση τιμή και την τυπική απόκλιση για κάθε ποσοτική μεταβλητή: $\bar{x} = 67,97$, $\bar{y} = 63,1$, $s_x = 10,93$, $s_y = 9,16$ και το άθροισμα

σμα $\sum_{i=1}^{30} x_i y_i = 131611$. Οπότε, $r=0,98$. Αυτό σημαίνει ότι το προσδόκιμο ζωής των γυναικών και των ανδρών στο δείγμα μας είναι ισχυρά θετικά γραμμικά συσχετισμένα.

- γ) Τα σημεία $(x_i, y_i), i = 1, 2, \dots, 30$ είναι συγκεντρωμένα γύρω από μια ευθεία, όπως φαίνεται στο σχήμα 2. Η εξίσωση της ευθείας δίνεται από τη σχέση (2). Επιλέγουμε δύο σημεία, έστω τα $A(45,44)$ και $B(81,74)$ πάνω στην ευθεία που φέραμε «με το μάτι». Αντικαθιστούμε τις συντεταγμένες (x,y) των σημείων αυτών στην (2), οπότε προκύπτει το σύστημα:

$$\begin{cases} y_A = \alpha + \beta x_A \\ y_B = \alpha + \beta x_B \end{cases} \Leftrightarrow \begin{cases} 44 = \alpha + 45\beta \\ 74 = \alpha + 81\beta \end{cases}$$

- Επιλύοντας το σύστημα αυτό βρίσκουμε $\alpha = -173,28$ και $\beta = 1,39$, οπότε η εξίσωση της ευθείας (1) γίνεται:

$$y = 6,5 + 0,83x$$

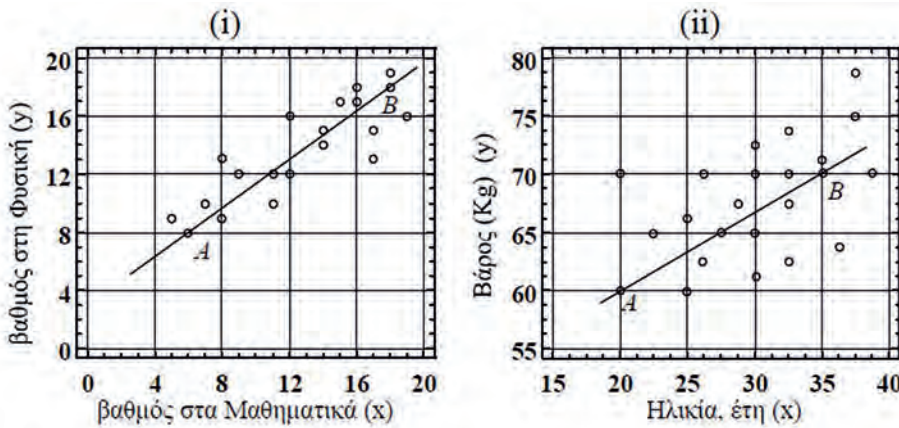
Επομένως, η ευθεία που κατά τη γνώμη μας προσαρμόζεται καλύτερα στα σημεία του διαγράμματος διασποράς διέρχεται από το σημείο $(0, 6,5)$ και έχει συντελεστή διεύθυνσης $0,83$.

- δ) Αν για την Αλβανία, το προσδόκιμο ζωής των γυναικών είναι τα 75 έτη, χρησιμοποιώντας την παραπάνω ευθεία μπορούμε να εκτιμήσουμε το προσδόκιμο ζωής των ανδρών της Αλβανίας. Έτσι, $y=6,5+0,83 \cdot 75=68,75$, δηλαδή 69 περίπου έτη.

Η θετική συσχέτιση μεταξύ Π.Ζ.Γ και Π.Ζ.Α. δεν μπορεί να αποδοθεί σε αιτιακή σχέση μεταξύ των Π.Ζ.Γ και Π.Ζ.Α., αλλά στην ύπαρξη συγχυτικών παραγόντων όπως, το επίπεδο διαβίωσης, το ΑΕΠ, την ποιότητα των υπηρεσιών υγείας κάθε χώρας, κτλ., που επηρεάζουν και τα δύο προσδόκιμα ζωής.

Ασκήσεις - Προβλήματα - Δραστηριότητες

- 1) Να διατάξετε τις παρακάτω τιμές του r σε αύξουσα τάξη του βαθμού γραμμικής συσχέτισης δύο ποσοτικών μεταβλητών X και Y : $-0,6, 0,9, -0,7, 0,2, 0, -1$.
- 2) Ο συντελεστής γραμμικής συσχέτισης δύο μεταβλητών X και Y είναι $0,96$, ενώ ο συντελεστής γραμμικής συσχέτισης δύο μεταβλητών Z και Φ είναι $-0,96$. Ποια είναι η διαφορά τους;
- 3) Δίνονται δυο διαγράμματα διασποράς με χαραγμένες «με το μάτι» δύο ευθείες από έναν μαθητή.



α) Χρησιμοποιώντας τα σημεία A και B να βρείτε τις εξισώσεις των δύο ευθειών.

β) Πώς θα μπορούσατε να χρησιμοποιήσετε τις ευθείες του ερωτήματος α);

4) Μια εταιρεία διαφημίσεων παρουσίασε τον επόμενο πίνακα:

Αριθμός διαφημίσεων	Έσοδα από πωλήσεις
10	20000
18	28000
24	35000
32	44000
35	48000
37	50000
42	55000

α) Να κατασκευάσετε το διάγραμμα διασποράς και να εκτιμήσετε από αυτό τον συντελεστή γραμμικής συσχέτισης μεταξύ του αριθμού διαφημίσεων της εταιρείας και των εσόδων της από τις πωλήσεις.

β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ του αριθμού διαφημίσεων της εταιρείας και των εσόδων της από τις πωλήσεις.

γ) Να σχεδιάσετε «με το μάτι» την ευθεία που φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα.

δ) Αν το πλήθος των διαφημίσεων που αγόραζε ένα κανάλι ήταν 30, πόσο εκτιμάτε ότι θα ήταν τα έσοδα της εταιρείας διαφημίσεων;

ε) Αν το πλήθος των διαφημίσεων που αγόραζε ένα κανάλι ήταν 60, θα ήταν ασφαλές να εκτιμήσετε τα έσοδα της εταιρείας διαφημίσεων, όπως στο δ);

- 5) Τα δεδομένα του επόμενου πίνακα παριστάνουν τους βαθμούς (στην κλίμακα του 100) 10 μαθητών/τριών της Β΄ τάξης του Γενικού Λυκείου στα μαθήματα της Φυσικής (X) και των Μαθηματικών (Y) κορμού σε μια γραπτή αξιολόγηση.

Βαθμός- X	Βαθμός- Y	Βαθμός- X	Βαθμός- Y
67	63	81	85
74	67	93	89
67	70	81	89
78	74	96	96
89	81	89	100

- α) Να κατασκευάσετε το διάγραμμα διασποράς και να εκτιμήσετε από αυτό τον συντελεστή γραμμικής συσχέτισης μεταξύ των βαθμών της Φυσικής (X) και των Μαθηματικών (Y) των 10 μαθητών/τριών της Β΄ τάξης του Γενικού Λυκείου.
- β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των βαθμών της Φυσικής (X) και των Μαθηματικών (Y) των 10 μαθητών/τριών της Β΄ τάξης του Γενικού Λυκείου.
- γ) Να σχεδιάσετε «με το μάτι» την ευθεία που φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα.
- δ) Πώς θα μπορούσατε να εκτιμήσετε τον βαθμό των Μαθηματικών ενός μαθητή της Β΄ Λυκείου, εάν γνωρίζατε ότι στη Φυσική έγραψε 70;
- 6) Στον παρακάτω πίνακα δίνονται οι ηλικίες και οι (συστολικές) πιέσεις αίματος 10 γυναικών.

Ηλικία (x)	56	42	72	36	63	47	55	49	38	60
Πίεση αίματος (y)	17	12	14	10	13	9	11	8	11	15

- α) Να σημειώσετε σε ορθοκανονικό σύστημα αξόνων τα σημεία (x, y) σύμφωνα με τον παραπάνω πίνακα, όπου x είναι η ηλικία των γυναικών σε έτη και y είναι η πίεση αίματος των γυναικών σε cm Hg.
- β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των μεταβλητών της ηλικίας των γυναικών σε έτη (x) και της πίεσης τους σε cm Hg (y).
- 7) Στον ακόλουθο πίνακα παρουσιάζονται στην πρώτη γραμμή οι τιμές (σε €) για διαφορετικά κράνη ποδηλασίας και στη δεύτερη γραμμή η βαθμολογία ποιότητάς τους που έγινε από ειδικούς (σε μια κλίμακα από 0 έως 100, όπου όσο μεγαλύτερη είναι η τιμή τόσο πιο ποιοτικό είναι το κράνος).

Τιμή (€)	35	22	33	42	50	23	29	18	39	28	20	25
Βαθμολογία ποιότητας	64	60	58	55	54	45	47	43	42	41	40	32

- α) Να κατασκευάσετε το διάγραμμα διασποράς.
- β) Υπάρχει γραμμική συσχέτιση ανάμεσα στην τιμή και τη βαθμολογία ποιότητας;
- γ) Θα μπορούσαμε να πούμε με βεβαιότητα ότι αν αγοράσουμε πιο φθηνό κράνος θα έχει πιο χαμηλή ποιότητα;
- δ) Να σχεδιάσετε «με το μάτι» στο διάγραμμα διασποράς μια ευθεία που θα μπορούσε να περιγράφει τη σχέση του αναμενόμενου βαθμού ποιότητας ενός ποδηλατικού κράνος με την τιμή του.
- 8) Στον ακόλουθο πίνακα παρουσιάζονται οι συντελεστές γραμμικής συσχέτισης των γραπτών βαθμολογιών στις εξετάσεις Ιουνίου σε 5 μαθήματα ενός τμήματος Β΄ τάξης γενικού λυκείου.

	Άλγεβρα	Βιολογία	Γλώσσα	Φυσική	Χημεία
Άλγεβρα	1,00				
Βιολογία	0,54	1,00			
Γλώσσα	0,76	0,81	1,00		
Φυσική	0,70	0,73	0,71	1,00	
Χημεία	0,41	0,80	0,67	0,66	1,00

Να εξετάσετε κατά πόσο υπάρχει, ισχυρή ή όχι, γραμμική συσχέτιση ανάμεσα στις βαθμολογίες στα 5 εξεταζόμενα μαθήματα των μαθητών/τριών αυτών.

- 9) Το πλήθος x των οχημάτων σε εκατομμύρια και ο αριθμός y των ατυχημάτων σε εκατοντάδες, σε 15 διαφορετικές χώρες, δίνονται από τον επόμενο πίνακα:

Χώρα	A	B	Γ	Δ	E	Z	H	Θ	I	K	Λ	M	N	Ξ	O
x	8,6	13,4	12,8	9,3	1,3	9,4	13,1	4,9	13,5	9,6	7,5	9,8	23,3	21	19,4
y	33	51	30	48	12	23	46	18	36	50	34	35	95	99	69

- α) Να κατασκευάσετε το διάγραμμα διασποράς μεταξύ των μεταβλητών του πλήθους των οχημάτων και του αριθμού των οχημάτων για τις 15 χώρες.
- β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των μεταβλητών του πλήθους των οχημάτων και του αριθμού των οχημάτων για τις 15 χώρες.

- 10) Από 8 γάμους που έγιναν σε μια εκκλησία ενός χωριού κατά τη διάρκεια ενός μηνός οι ηλικίες των ανδρόγυνων ήταν:

Ηλικία γαμπρού	20	22	24	25	28	30	33	38
Ηλικία νύφης	20	20	22	27	24	25	28	34

- α) Να κατασκευάσετε το διάγραμμα διασποράς μεταξύ των ηλικιών της νύφης (Y) και του γαμπρού (X) και να περιγράψετε το είδος της σχέσης που φαίνεται να έχουν οι δύο μεταβλητές.
- β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των ηλικιών νύφης και γαμπρού.
- γ) Να σχεδιάσετε «με το μάτι» την ευθεία που φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα.
- δ) Να βρείτε την αναμενόμενη ηλικία της νύφης για έναν υποψήφιο γαμπρό ετών 34.
- 11) Δίνεται δείγμα n ζευγών παρατηρήσεων (x_i, y_i) , $i=1,2,\dots,n$ δύο μεταβλητών X και Y και έστω r , ο συντελεστής γραμμικής συσχέτισης των μεταβλητών X και Y . Να αποδείξετε ότι αν όλα τα παραπάνω σημεία βρίσκονται πάνω σε μια ευθεία με εξίσωση:

$$y_i = \alpha + \beta x_i, \quad i = 1, 2, \dots, n$$

τότε:

$$r = 1 \text{ αν } \alpha > 0 \text{ και } r = -1 \text{ αν } \alpha < 0.$$

Πρόσθετο Υλικό

Θέματα για διερεύνηση



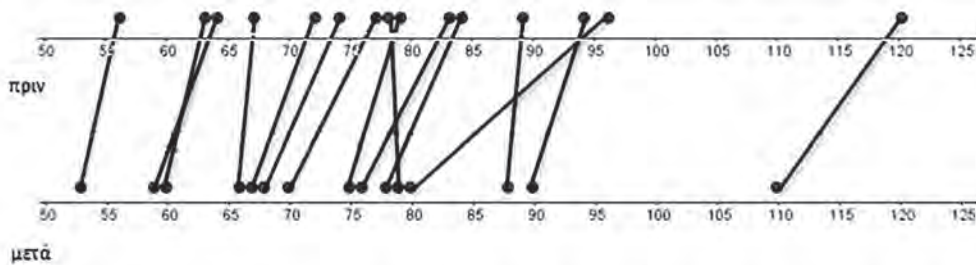
- 1) Σε μια έρευνα που έγινε με σκοπό να διερευνηθεί η αποτελεσματικότητα μιας δίαιτας, μετρήθηκε το βάρος 15 ατόμων πριν και μετά τη δίαιτα.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Πριν	89	83	78	56	96	120	67	63	72	74	79	94	84	64	77
Μετά	88	76	79	53	80	110	66	60	67	67	75	90	78	59	70

- α) Να υπολογίσετε τη μέση τιμή και την τυπική απόκλιση της μεταβλητής του βάρους των 15 ατόμων πριν ($X_{\text{πριν}}$) και μετά ($X_{\text{μετά}}$) τη δίαιτα. Να σχολιάσετε τα αποτελέσματα.
- β) Να κατασκευάσετε τα θηκογράμματα για τη μεταβλητή του βάρους των 15 ατόμων πριν και μετά τη δίαιτα. Να τα συγκρίνετε και να διατυπώσετε την

άποψή σας για το αν υπάρχει διαφορά πριν και μετά και τη δίαιτα.

- γ) Δίνονται τα επόμενα σημειογράμματα για τη μεταβλητή του βάρους των 15 ατόμων πριν και μετά τη δίαιτα, στα οποία έχουν ενωθεί τα βάρη των 15 ατόμων πριν και μετά τη δίαιτα. Τι παρατηρείτε από το γράφημα; Επιβεβαιώνεται η παρατήρησή σας, συγκρίνοντας με τα αποτελέσματα των προηγούμενων ερωτημάτων;



- δ) Να δημιουργήσετε τη μεταβλητή $Z = X_{\text{πριν}} - X_{\text{μετά}}$ για κάθε άτομο και να κατασκευάσετε το θηκόγραμμά της. Πώς μπορείτε να αναδείξετε τη διαφορά των τιμών του βάρους των 15 ατόμων πριν και μετά τη δίαιτα από αυτό το θηκόγραμμα;
- ε) Να κατασκευάσετε τα διαγράμματα διασποράς ανάμεσα στις μεταβλητές του βάρους ($X_{\text{πριν}}$) και ($X_{\text{μετά}}$) των 15 ατόμων.
- στ) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των μεταβλητών του βάρους ($X_{\text{πριν}}$) και ($X_{\text{μετά}}$) των 15 ατόμων.
- ζ) Να σχεδιάσετε «με το μάτι» την ευθεία που φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα.
- η) Θα μπορούσατε να εκτιμήσετε το βάρος ενός ατόμου που πρόκειται να ακολουθήσει αυτή τη δίαιτα, εάν το αρχικό του βάρος ήταν 91 κιλά;
- 2) Επισκεφθείτε την ιστοσελίδα της ΕΛ.ΣΤΑΤ. (www.statistics.gr) και επιλέξτε ένα κατάλληλο σύνολο δεδομένων που να περιέχει τουλάχιστον δύο ποσοτικές μεταβλητές. Αποφασίστε ποια θα είναι η εξαρτημένη και ποια η ανεξάρτητη μεταβλητή. Στη συνέχεια, κάντε το διάγραμμα διασποράς των δύο μεταβλητών, υπολογίστε και ερμηνεύστε τον συντελεστή γραμμικής συσχέτισης και χαράξτε με το μάτι μια ευθεία που φαίνεται να προσαρμόζεται καλά στα δεδομένα σας. Τέλος, κάντε μια-δυο εκτιμήσεις για διάφορες τιμές της ανεξαρτητής σας μεταβλητής που να ανήκουν στο εύρος τιμών της. Καταγράψτε τα αποτελέσματά σας, συντάσσοντας μια μικρή αναφορά (μέχρι 2 σελίδες, με πίνακες και γραφήματα).
- 3) Να ετοιμάσετε ένα ερωτηματολόγιο για να μελετήσετε ορισμένα χαρακτηριστικά από την καθημερινότητά σας και να τα συσχετίσετε με τον γενικό βαθμό που πήραν κατά το τελευταίο τετράμηνο. Τα χαρακτηριστικά που μπορείτε να



καταγράψετε είναι: πόσες ώρες διαβάζετε καθημερινά κατά μέσο όρο, πόσες εξωσχολικές δραστηριότητες έχετε, πόσες ώρες κατά μέσο όρο ξοδεύετε σε αυτές και ό,τι άλλο θα επιθυμούσατε εσείς. Αναλύστε τα δεδομένα με αυτά που μάθατε σε αυτήν την παράγραφο. Περιγράψτε τη σχέση που έχει καθένα από αυτά τα χαρακτηριστικά με τον βαθμό του τετραμήνου. Καταγράψτε τα αποτελέσματά σας, συντάσσοντας μια μικρή αναφορά (μέχρι 5 σελίδες, με πίνακες και γραφήματα).