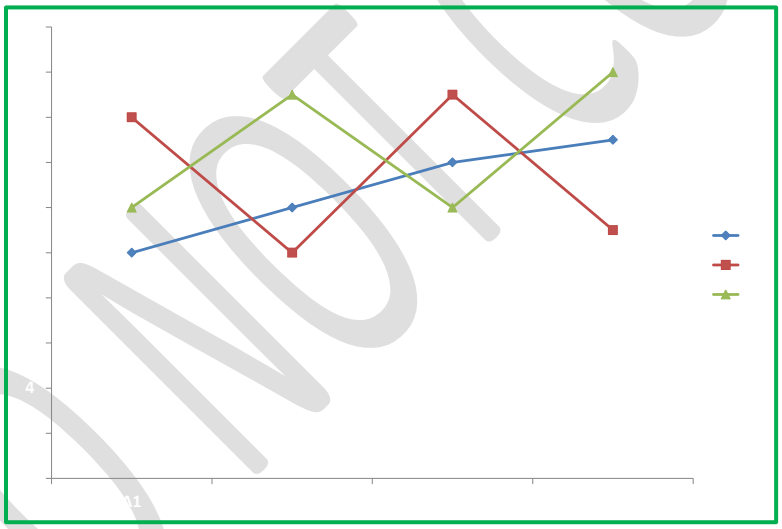
# Τμήμα Γεωπονίας ΑΠΘ

# Εργαστήριο “Γεωργίας”

# Τετράδιο Ασκήσεων Βιομετρίας (ΙΙ)




## Γεώργιος Μενεξές

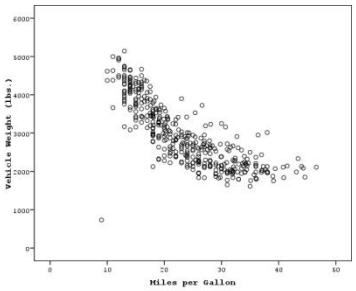
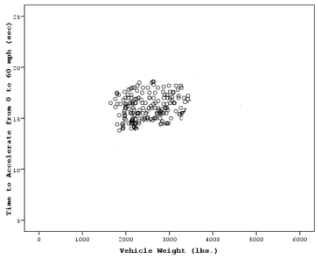
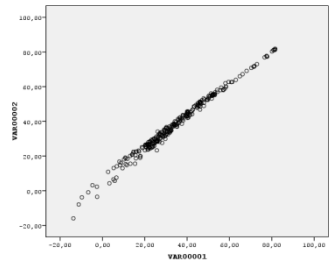
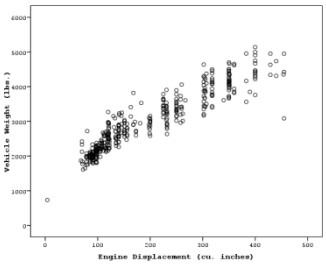
**Αναπληρωτής Καθηγητής**


**Θεσσαλονίκη, Μάιος 2021**

DO NOT COPY

## Άσκηση 1

Κοιτάξτε με προσοχή τα παρακάτω διαγράμματα (Α-Δ) διασποράς-συμμεταβολής δύο ποσοτικών μεταβλητών.

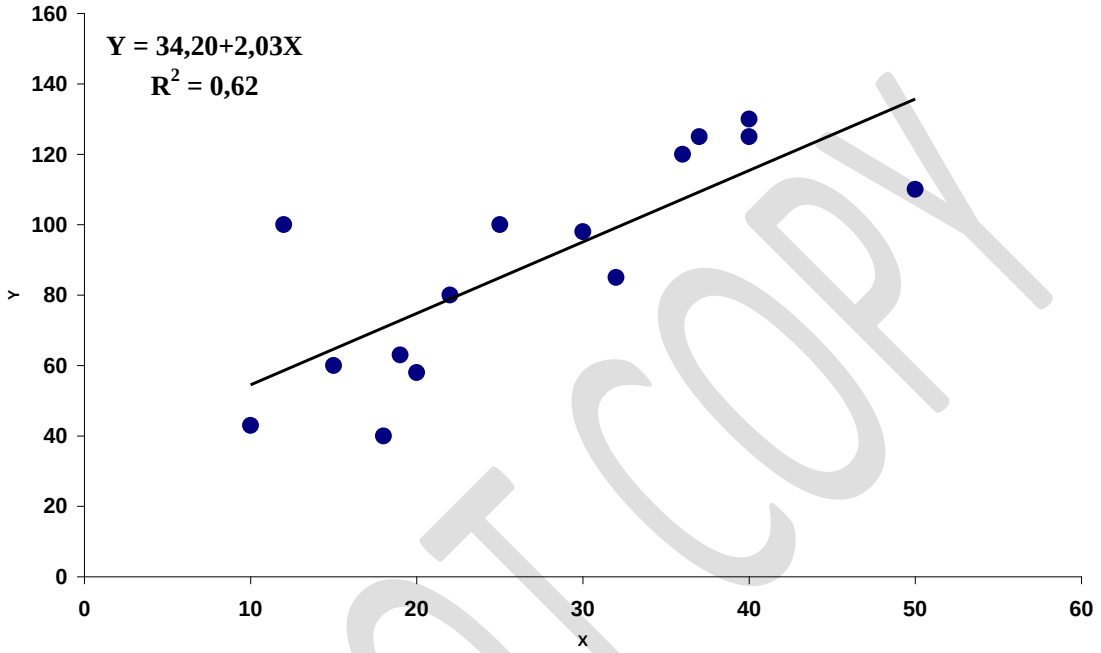
| | |
|---|---|
|   |   |
| **Α** | **Β** |
|   |   |
| **Γ** | **Δ** |

1. Σε ποιο ή ποια διαγράμματα φαίνεται να υπάρχει **θετική συσχέτιση** μεταξύ των δύο μεταβλητών που εξετάζονται;
2. Σε ποιο ή ποια διαγράμματα φαίνεται να υπάρχει **αρνητική συσχέτιση** μεταξύ των δύο μεταβλητών που εξετάζονται;
3. Σε ποιο ή ποια διαγράμματα φαίνεται να υπάρχει **γραμμική συσχέτιση** μεταξύ των δύο μεταβλητών που εξετάζονται;
4. Σε ποιο ή ποια διαγράμματα φαίνεται να υπάρχει **καμπυλόγραμμη συσχέτιση** μεταξύ των δύο μεταβλητών που εξετάζονται;
5. Σε ποιο ή ποια διαγράμματα φαίνεται **να μην υπάρχει συσχέτιση** μεταξύ των δύο μεταβλητών που εξετάζονται;

DO NOT COPY

## Άσκηση 2

Κοιτάξτε με προσοχή το παρακάτω διαγράμματα διασποράς-συμμεταβολής δύο ποσοτικών μεταβλητών X (kg λιπάσματος) και Y (απόδοση καλλιέργειας), την εξίσωση απλής ευθύγραμμης συμμεταβολής και την ευθεία ελαχίστων τετραγώνων.




Στην εξίσωση $Y=a+bX$

1. Ποια είναι η ανεξάρτητη μεταβλητή (στη συγκεκριμένη περίπτωση);
2. Ποια είναι η εξαρτημένη μεταβλητή (στη συγκεκριμένη περίπτωση);
3. Ποια είναι η φορά (κατεύθυνση) της συσχέτισης/συμμεταβολής μεταξύ των δύο μεταβλητών;
4. Τι εκφράζει ο συντελεστής $a$ (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση);
5. Τι εκφράζει ο συντελεστής $b$ (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση);
6. Τι εκφράζει ο συντελεστής προσδιορισμού $R^2$ (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση));
7. Από την τιμή του $R^2$ να υπολογίσετε την τιμή του συντελεστή γραμμικής συσχέτισης $r$ του Pearson μεταξύ των δύο μεταβλητών (απόδοση, ποσότητα λιπάσματος).
8. Ποια είναι η εκτίμηση της απόδοσης $Y$ για $X$=23 kg λιπάσματος;
9. Με βάση την τιμή του $R^2$ είναι καλή η εκτίμηση;
10. Πάνω στο διάγραμμα διασποράς εντοπίστε τα σημεία με το μεγαλύτερο σφάλμα.

DO NOT COPY

## Άσκηση 3

Δεκαπέντε φυτά φράουλας καλλιεργήθηκαν στο θερμοκήπιο. Κατά την περίοδο της συγκομιδής ελήφθη η απόδοση σε γραμμάρια καθώς και η περιεκτικότητα των φύλων σε άζωτο. Τα αποτελέσματα δίνονται στον Πίνακα 1 που ακολουθεί.

**Πίνακας 1**

| Απόδοση (Y) | Άζωτο (X) |
|---|---|
| 247 | 2,50 |
| 245 | 2,55 |
| 266 | 2,54 |
| 277 | 2,56 |
| 284 | 2,68 |
| 251 | 2,57 |
| 275 | 2,62 |
| 272 | 2,58 |
| 241 | 2,63 |
| 265 | 2,59 |
| 281 | 2,69 |
| 292 | 2,61 |
| 285 | 2,67 |
| 274 | 2,52 |
| 282 | 2,53 |

1. Να υπολογίσετε το μέσο όρο της απόδοσης και το μέσο όρο της περιεκτικότητας σε άζωτο.
2. Να σχεδιάσετε το διάγραμμα διασποράς των δύο μεταβλητών (Άζωτο, Απόδοση).
3. Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης $r$ του *Pearson* μεταξύ των δύο μεταβλητών.
4. Διατυπώστε μια υπόθεση βάσει της οποίας είναι δυνατό οι δύο μεταβλητές να συνδεθούν με σχέση αιτίας-αποτελέσματος.
5. Να σχεδιάσετε την αντίστοιχη Ευθεία Ελαχίστων Τετραγώνων (Ευθύγραμμης Συμμεταβολής) πάνω στο διάγραμμα διασποράς των δύο μεταβλητών.
6. Να ερμηνεύσετε τους συντελεστές της εξίσωσης παλινδρόμησης (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση).
7. Να υπολογίσετε και να ερμηνεύσετε το συντελεστή προσδιορισμού $R^2$ (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση).
8. Να εκτιμήσετε την απόδοση σε γραμμάρια όταν η αντίστοιχη περιεκτικότητα σε άζωτο είναι 2,62.
9. Είναι καλή η εκτίμηση με βάση την τιμή του δείκτη $R^2$;
10. Πάνω στο διάγραμμα διασποράς εντοπίστε τα σημεία με το μεγαλύτερο σφάλμα.

DO NOT COPY

## Άσκηση 4

Ένας ερευνητής μελέτησε την ποσότητα οξυγόνου ($mm^3$) που καταναλώνουν μαρίδες σε σχέση με το βάρος τους (gr). Τα δεδομένα δίνονται στον Πίνακα 2 που ακολουθεί.

**Πίνακας 2**

| Βάρος Μαρίδων | Ποσότητα Οξυγόνου ($mm^3$) |
|---|---|
| 7 | 188 |
| 6 | 170 |
| 8 | 249 |
| 6 | 121 |
| 7 | 201 |
| 8 | 269 |
| 8 | 239 |
| 4 | 130 |
| 5 | 107 |
| 6 | 180 |
| 7 | 142 |
| 10 | 270 |
| 7 | 249 |
| 5 | 130 |
| 4 | 101 |

1. Να υπολογίσετε το μέσο όρο, το τυπικό σφάλμα στην εκτίμηση του μέσου όρου, την τυπική απόκλιση και τον συντελεστή *CV* για κάθε μία μεταβλητή (βάρος, ποσότητα Οξυγόνου).
2. Τι παρατηρείτε για την πολλαπλότητα των τιμών του βάρους των μαρίδων;
3. Να σχεδιάσετε το διάγραμμα διασποράς των δύο μεταβλητών (βάρος, ποσότητα Οξυγόνου).
4. Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης $r$ του *Pearson* μεταξύ των δύο μεταβλητών.
5. Διατυπώστε μια υπόθεση βάσει της οποίας είναι δυνατό οι δύο μεταβλητές να συνδεθούν με σχέση αιτίας-αποτελέσματος.
6. Να σχεδιάσετε την αντίστοιχη ευθεία ελαχίστων τετραγώνων πάνω στο διάγραμμα διασποράς των δύο μεταβλητών.
7. Να γράψετε το μαθηματικό υπόδειγμα της ευθείας ελαχίστων τετραγώνων.
8. Να ερμηνεύσετε τους συντελεστές της εξίσωσης παλινδρόμησης (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση).
9. Να βρείτε δύο σημεία του επιπέδου από τα οποία διέρχεται η ευθεία ελαχίστων τετραγώνων.
10. Να υπολογίσετε και να ερμηνεύσετε το συντελεστή προσδιορισμού $R^2$ (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση).
11. Προσαρμόζονται καλά τα δεδομένα στο υπόδειγμα της γραμμικής παλινδρόμησης; Οι αποκλίσεις είναι στατιστικά σημαντικές, σε επίπεδο σημαντικότητας 0,05, ή όχι;
12. Να εκτιμήσετε την ποσότητα οξυγόνου που καταναλώνουν μαρίδες με βάρος 9gr.
13. Είναι καλή η εκτίμηση με βάση την τιμή του δείκτη $R^2$;

DO NOT COPY

## Άσκηση 5

Ένα λίπασμα χρησιμοποιήθηκε στα ζαχαρότευτλα σε 6 διαφορετικές δόσεις και ελήφθησαν οι αντίστοιχες αποδόσεις. Οι μετρήσεις της απόδοσης είναι μέσοι όροι από 4 επαναλήψεις (πειραματικά τεμάχια). Τα δεδομένα δίνονται στον Πίνακα 3.

**Πίνακας 3**

| X: Ποσότητα λιπάσματος | Y: Απόδοση σε Kg |
|---|---|
| 0,5 | 10 |
| 1 | 16 |
| 2 | 26 |
| 3 | 35 |
| 4 | 50 |
| 6 | 72 |

1. Να σχεδιάσετε το διάγραμμα διασποράς των δύο μεταβλητών (Ποσότητα, Απόδοση).
2. Να υπολογίσετε και να ερμηνεύσετε το συντελεστή γραμμικής συσχέτισης $r$ του *Pearson* μεταξύ των δύο μεταβλητών.
3. Είναι στατιστικά σημαντικός σε επίπεδο σημαντικότητας $\alpha$=0,05;
4. Το αποτέλεσμα/συμπέρασμα του ελέγχου σημαντικότητας του συντελεστή συσχέτισης είναι ασφαλές;
5. Διατυπώστε μια υπόθεση βάσει της οποίας είναι δυνατό οι δύο μεταβλητές να συνδεθούν με σχέση αιτίας-αποτελέσματος.
6. Να σχεδιάσετε την αντίστοιχη Ευθεία Ελαχίστων Τετραγώνων πάνω στο διάγραμμα διασποράς των δύο μεταβλητών.
7. Να ερμηνεύσετε τους συντελεστές της εξίσωσης παλινδρόμησης (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση).
8. Να υπολογίσετε και να ερμηνεύσετε το συντελεστή προσδιορισμού $R^2$ (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση).
9. Να εκτιμήσετε την απόδοση όταν η αντίστοιχη δόση λιπάσματος είναι 2,5.
10. Είναι καλή η εκτίμηση με βάση την τιμή του δείκτη $R^2$;
11. Τι εκφράζει η ποσότητα/δείκτης $e = b.\frac{\bar{X}}{\bar{Y}}$, όπου $b$ είναι η κλίση της ευθείας ελαχίστων τετραγώνων και $\bar{X}$ και $\bar{Y}$ είναι οι μέσοι όροι της ποσότητας λιπάσματος και της απόδοσης αντίστοιχα. Να υπολογιστεί και να ερμηνευτεί ο δείκτης $e$ στη συγκεκριμένη περίπτωση.
    <u>Υπόδειξη:</u> Ο δείκτης $e$ ονομάζεται συντελεστής ελαστικότητας της εξαρτημένης μεταβλητής ως προς την ανεξάρτητη μεταβλητή.

DO NOT COPY

## Άσκηση 6

Στον Πίνακα 4 που ακολουθεί δίνεται η απόδοση (%) ενός φυτικού είδους (Y) και ο αντίστοιχος βαθμός ωριμότητάς του (X). Οι μετρήσεις του βαθμού ωριμότητάς είναι μέσοι όροι από 4 επαναλήψεις.

**Πίνακας 4**

| Βαθμός Ωριμότητας | Απόδοση (%) |
|---|---|
| 11 | 34 |
| 14 | 28 |
| 15 | 30 |
| 16 | 32 |
| 17 | 25 |
| 18 | 23 |

1. Να σχεδιάσετε το διάγραμμα διασποράς των δύο μεταβλητών (Ωριμότητα, Απόδοση).
2. Να υπολογίσετε και να ερμηνεύσετε το συντελεστή γραμμικής συσχέτισης $r$ του *Pearson* μεταξύ των δύο μεταβλητών.
3. Είναι στατιστικά σημαντικός σε επίπεδο σημαντικότητας $\alpha$=0,05;
4. Το αποτέλεσμα/συμπέρασμα του ελέγχου σημαντικότητας του συντελεστή συσχέτισης είναι ασφαλές;
5. Διατυπώστε μια υπόθεση βάσει της οποίας είναι δυνατό οι δύο μεταβλητές (Ωριμότητα, Απόδοση) να συνδεθούν με σχέση αιτίας-αποτελέσματος.
6. Να σχεδιάσετε την αντίστοιχη Ευθεία Ελαχίστων Τετραγώνων (Ευθύγραμμης Συμμεταβολής) πάνω στο διάγραμμα διασποράς των δύο μεταβλητών.
7. Να ερμηνεύσετε τους συντελεστές της εξίσωσης παλινδρόμησης (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση).
8. Να υπολογίσετε και να ερμηνεύσετε το συντελεστή προσδιορισμού $R^2$ (όχι γενική απάντηση αλλά στη συγκεκριμένη περίπτωση).
9. Να εκτιμήσετε την απόδοση όταν ο βαθμός ωριμότητας είναι 12.
10. Είναι "καλή" η εκτίμηση με βάση την τιμή του δείκτη $R^2$;
11. Τι εκφράζει η ποσότητα/δείκτης $e = b.\frac{\bar{X}}{\bar{Y}}$, όπου $b$ είναι η κλίση της ευθείας ελαχίστων τετραγώνων και $\bar{X}$ και $\bar{Y}$ είναι οι μέσοι όροι του βαθμού ωριμότητας και της απόδοσης αντίστοιχα. Να υπολογιστεί και να ερμηνευτεί ο δείκτης $e$ στη συγκεκριμένη περίπτωση.
    <u>Υπόδειξη:</u> Ο δείκτης $e$ ονομάζεται συντελεστής ελαστικότητας της εξαρτημένης μεταβλητής ως προς την ανεξάρτητη μεταβλητή.

DO NOT COPY

## Άσκηση 8

Στον Πίνακα 5 δίνονται 100 ζεύγη τιμών δύο μεταβλητών X1 και X2 (τα δεδομένα ελήφθησαν με τυχαία δειγματοληψία).

**Πίνακας 5**

| X1 | X2 | X1 | X2 | X1 | X2 |
|---|---|---|---|---|---|
| 102,503 | 141,502 | 38,229 | 76,724 | 96,519 | 118,690 |
| 110,307 | 143,387 | 42,433 | 94,397 | 157,238 | 202,060 |
| 120,286 | 176,975 | 104,050 | 158,508 | 68,864 | 113,481 |
| 96,054 | 121,390 | 80,676 | 152,527 | 86,040 | 143,717 |
| 124,905 | 182,333 | 64,143 | 136,691 | 86,460 | 134,832 |
| 98,256 | 123,819 | 81,179 | 115,111 | 102,973 | 127,043 |
| 75,807 | 83,511 | 143,280 | 222,891 | 133,977 | 148,506 |
| 79,685 | 115,520 | 116,876 | 155,240 | 107,080 | 151,728 |
| 137,929 | 209,379 | 84,849 | 160,399 | 71,207 | 137,035 |
| 90,418 | 146,677 | 79,028 | 191,042 | 69,888 | 115,681 |
| 92,357 | 170,400 | 96,355 | 131,857 | 90,878 | 140,613 |
| 74,398 | 124,928 | 104,440 | 124,147 | 95,827 | 157,034 |
| 197,920 | 249,868 | 103,659 | 95,089 | 67,120 | 129,693 |
| 163,800 | 178,416 | 131,926 | 162,956 | 110,103 | 151,323 |
| 115,078 | 145,903 | 126,699 | 146,538 | 111,650 | 150,707 |
| 146,891 | 178,773 | 87,582 | 123,339 | 72,276 | 142,462 |
| 104,649 | 182,442 | 96,290 | 162,016 | 125,770 | 136,026 |
| 128,906 | 154,749 | 66,472 | 106,787 | 125,447 | 134,954 |
| 106,360 | 122,592 | 57,938 | 99,204 | 76,820 | 122,268 |
| 98,339 | 133,358 | 88,526 | 94,063 | 93,401 | 109,155 |
| 92,296 | 144,674 | 111,952 | 154,537 | 132,481 | 165,246 |
| 143,009 | 173,735 | 109,936 | 152,187 | 121,690 | 187,267 |
| 55,728 | 91,931 | 92,729 | 102,281 | 73,735 | 135,421 |
| 104,454 | 180,461 | 133,225 | 174,195 | 65,365 | 152,150 |
| 90,485 | 109,601 | 129,567 | 226,324 | 93,863 | 161,522 |
| 151,091 | 151,230 | 98,720 | 169,702 | 117,878 | 123,095 |
| 111,637 | 122,114 | 87,014 | 108,961 | 132,855 | 213,129 |
| 118,056 | 153,201 | 161,771 | 153,755 | 84,131 | 146,528 |
| 117,738 | 162,689 | 87,744 | 99,964 | 154,486 | 198,402 |
| 70,980 | 144,650 | 52,460 | 113,091 | 96,978 | 122,496 |
| 152,699 | 172,132 | 120,419 | 208,993 | 167,062 | 205,624 |
| 157,757 | 173,440 | 59,137 | 122,344 | 76,958 | 110,190 |
| 160,649 | 207,913 | 94,413 | 136,269 | | |
| 126,570 | 148,540 | 78,092 | 139,819 | | |

DO NOT COPY

1. Να κατασκευάσετε το διάγραμμα διασποράς των δύο μεταβλητών (πχ το X1×X2).
2. Από το διάγραμμα διασποράς, να κάνετε μια εμπειρική εκτίμηση για την κατεύθυνση, το είδος και την ένταση της συσχέτισης μεταξύ των δύο μεταβλητών.
3. Να κάνετε έλεγχο σχετικά με το εάν υπάρχουν μονομεταβλητά και πολυμεταβλητά (2D) παράτυπα σημεία (outliers).
4. Να κάνετε έλεγχο αν οι δύο μεταβλητές ακολουθούν η κάθε μία Κανονική Κατανομή και να εκτιμήσετε τις αντίστοιχες παραμέτρους.
5. Να κάνετε έλεγχο αν οι δύο μεταβλητές ακολουθούν τη Διδιάστατη Κανονική Κατανομή και να εκτιμήσετε τις παραμέτρους αυτής.
6. Να υπολογίσετε τη συνδυασπορά (covariance) των δύο μεταβλητών.
7. Να υπολογίσετε τον συντελεστή γραμμικής συσχέτισης (*r* του Pearson) μεταξύ των δύο μεταβλητών.
8. Πως συνδέεται ο συντελεστής συσχέτισης των δύο μεταβλητών με τη συνδυασπορά τους;
9. Να αξιολογήσετε τη στατιστική σημαντικότητα και την ένταση/βαθμό του συντελεστή συσχέτισης.
   Υπόδειξη: Η στατιστική σημαντικότητα του συντελεστή συσχέτισης να αξιολογηθεί με παραμετρικό έλεγχο και με τις μεθόδους προσομοίωσης Monte-Carlo και Bootstrap.
10. Για κάθε μεταβλητή να υπολογίσετε τα αντίστοιχα z-scores.
11. Πως συνδέονται τα z-scores με το συντελεστή γραμμικής συσχέτισης των δύο μεταβλητών;
12. Να εκτιμήσετε την ευθεία παλινδρόμησης της X1 πάνω στη X2.
    Υπόδειξη: Να γίνει πλήρης στατιστική ανάλυση, παρουσίαση, σχολιασμός και ερμηνεία των αποτελεσμάτων καθώς και έλεγχος των προϋποθέσεων.
13. Με τη μέθοδο προσομοίωσης Bootstrap να εκτιμήσετε 95% διαστήματα εμπιστοσύνης για τους συντελεστές του υποδείγματος (a και b).
14. Να κατασκευάσετε το κατάλληλο διάγραμμα διασποράς.
15. Επί του διαγράμματος διασποράς να σχεδιάσετε την ευθεία ελαχίστων τετραγώνων καθώς και την καμπύλη Loess (30%, 50% και 70% των γειτονικών σημείων).
    Υπόδειξη: Να κατασκευαστούν 3 διαγράμματα διασποράς ανάλογα με το ποσοστό των γειτονικών σημείων βάσει των οποίων θα σχεδιαστεί η καμπύλη Loess.
16. Τι παρατηρείτε;
17. Να εκτιμήσετε την τιμή της εξαρτημένης μεταβλητής όταν η ανεξάρτητη μεταβλητή πάρει τις τιμές 262, 290 και 145. Να υπολογίσετε και τα αντίστοιχα σφάλματα στις εκτιμήσεις.
18. Να εκτιμήσετε την ευθεία παλινδρόμησης της X2 πάνω στη X1.
    Υπόδειξη: Να γίνει πλήρης στατιστική ανάλυση, παρουσίαση, σχολιασμός και ερμηνεία των αποτελεσμάτων καθώς και έλεγχος των προϋποθέσεων.
19. Με τη μέθοδο προσομοίωσης Bootstrap να εκτιμήσετε 95% διαστήματα εμπιστοσύνης για τους συντελεστές του υποδείγματος (a και b).
20. Να κατασκευάσετε το κατάλληλο διάγραμμα διασποράς.
21. Επί του διαγράμματος διασποράς να σχεδιάσετε την ευθεία ελαχίστων τετραγώνων καθώς και την καμπύλη Loess (30%, 50% και 70% των γειτονικών σημείων).
    Υπόδειξη: Να κατασκευαστούν 3 διαγράμματα διασποράς ανάλογα με το ποσοστό των γειτονικών σημείων βάσει των οποίων θα σχεδιαστεί η καμπύλη Loess.

DO NOT COPY

22. Τι παρατηρείτε;
23. Να διατυπώσετε μία τουλάχιστον γεωμετρική ερμηνεία του συντελεστή γραμμικής συσχέτισης του Pearson.
24. Να υπολογίσετε τον συντελεστή συσχέτισης των τάξεων (*rho* του Spearman) μεταξύ των δύο μεταβλητών.
25. Να αξιολογήσετε τη στατιστική σημαντικότητα και την ένταση/βαθμό του συντελεστή συσχέτισης των τάξεων.
26. Τι παρατηρείτε για τις τιμές των δύο συντελεστών συσχέτισης *r* και *rho*;
27. Να επιλέξετε τυχαία 30 ζεύγη τιμών για τις μεταβλητές X1 και X2 και να απαντήσετε στα προηγούμενα ερωτήματα.
28. Να συγκρίνετε τα ευρήματα που προέκυψαν από την ανάλυση των 100 ζευγών με αυτά που προέκυψαν από την ανάλυση των 30 τυχαίων ζευγών.

DO NOT COPY

## Άσκηση 9

Η ετήσια συνολική κατανάλωση ενός προϊόντος (Y σε τόνους) σε μια μικρή χώρα για τα έτη 2011-2019 παρουσιάζεται στον Πίνακα 6. Θέλουμε να προσαρμόσουμε τα δεδομένα στο μαθηματικό υπόδειγμα:

$$Y_t=\gamma\alpha^{(t-2000)}$$

1. Να γίνει γραφική παράσταση των δεδομένων (διάγραμμα διασποράς- scatter plot). Επί του διαγράμματος διασποράς να σχεδιάσετε την καμπύλη Loess (40% των γειτονικών σημείων).
2. Να υπολογίσετε τους εκτιμητές ελαχίστων τετραγώνων των συντελεστών α και γ.
3. Να πραγματοποιήσετε όλους τους σχετικούς ελέγχους των προϋποθέσεων εγκυρότητας του υποδείγματος.
4. Να κάνετε έλεγχο και αν τα υπόλοιπα (residuals) του υποδείγματος έχουν αυτοσυσχέτιση.
5. Στο διάγραμμα διασποράς των δύο μεταβλητών (t, $Y_t$) να σχεδιάσετε την καμπύλη του υποδείγματος.
6. Στο διάγραμμα του ερωτήματος 5 να σχεδιάσετε και την καμπύλη των κινητών μέσων όρων 5 περιόδων.
7. Προσαρμόζονται καλά τα δεδομένα στο υπόδειγμα; Αιτιολογείστε την απάντησή σας με βάση αριθμητικά και διαγραμματικά αποτελέσματα.
8. Να γίνει πρόβλεψη της κατανάλωσης για τα έτη 2020 και 2021.
9. Να γίνει εκτίμηση των σφαλμάτων για τις δύο προβλέψεις του ερωτήματος 8.
10. Με βάση το αρχικό προτεινόμενο υπόδειγμα να εκτιμήσετε την κατανάλωση για την δεκαετία 2022-2031.
11. Με τα νέα δεδομένα να απαντήσετε στα ερωτήματα 1-8 τόσο για την περίοδο 2022-2031 όσο και για την περίοδο 2011-2031.

**Πίνακας 6**

| **Έτος (t)** | **Κατανάλωση ($Y_t$)** |
|---|---|
| 2011 | 97 |
| 2012 | 110 |
| 2013 | 147 |
| 2014 | 180 |
| 2015 | 230 |
| 2016 | 269 |
| 2017 | 406 |
| 2018 | 485 |
| 2019 | 615 |

DO NOT COPY

## Άσκηση 10

Ένας γεωργός θέλει να μελετήσει πως η ποσότητα λιπάσματος (X) σε μια συγκεκριμένη καλλιέργεια επηρεάζει την απόδοση (Y). Ο γεωργός διάλεξε 10 όσο το δυνατόν πιο όμοια αγροτεμάχια (με το ίδιο εμβαδό και στην ίδια περιοχή, ώστε να επικρατούν περίπου οι ίδιες εδαφοκλιματικές συνθήκες) και πειραματίστηκε με 10 διαφορετικές ποσότητες λιπάσματος. Στον Πίνακα 7 παρουσιάζονται τα αποτελέσματα του εγχειρήματος.

**Πίνακας 7**

| Αγροτεμάχιο | X: Ποσότητα Λιπάσματος (kg) | Y: Απόδοση (kg) |
|---|---|---|
| 1 | 20 | 710 |
| 2 | 10 | 560 |
| 3 | 26 | 793 |
| 4 | 8 | 527 |
| 5 | 19 | 674 |
| 6 | 16 | 614 |
| 7 | 17 | 735 |
| 8 | 12 | 581 |
| 9 | 9 | 539 |
| 10 | 24 | 759 |

1. Να γίνει πλήρης στατιστική ανάλυση, παρουσίαση, σχολιασμός και ερμηνεία των αποτελεσμάτων καθώς και έλεγχος των προϋποθέσεων.
2. Να γίνει εκτίμηση της απόδοσης για ποσότητες λιπάσματος 5, 15, και 30 kg και να υπολογιστούν τα σφάλματα των εκτιμήσεων. Ποιες διαπιστώσεις προκύπτουν από τις τιμές των σφαλμάτων των εκτιμήσεων;

DO NOT COPY

## Άσκηση 11

Στον Πίνακα 8 παρουσιάζονται τα αποτελέσματα ενός πειράματος (CRD, 5 επαναλήψεις) όπου εξετάστηκε η επίδραση 3 διαφορετικών λιπασμάτων στην απόδοση ενός φυτικού είδους.

**Πίνακας 8**

| Λίπασμα | Απόδοση | Επανάληψη | D1 | D2 | D3 |
|---|---|---|---|---|---|
| 1 | 750 | 1 | 1 | 0 | 0 |
| 1 | 740 | 2 | 1 | 0 | 0 |
| 1 | 755 | 3 | 1 | 0 | 0 |
| 1 | 745 | 4 | 1 | 0 | 0 |
| 1 | 730 | 5 | 1 | 0 | 0 |
| 2 | 690 | 1 | 0 | 1 | 0 |
| 2 | 650 | 2 | 0 | 1 | 0 |
| 2 | 630 | 3 | 0 | 1 | 0 |
| 2 | 610 | 4 | 0 | 1 | 0 |
| 2 | 660 | 5 | 0 | 1 | 0 |
| 3 | 450 | 1 | 0 | 0 | 1 |
| 3 | 490 | 2 | 0 | 0 | 1 |
| 3 | 500 | 3 | 0 | 0 | 1 |
| 3 | 520 | 4 | 0 | 0 | 1 |
| 3 | 430 | 5 | 0 | 0 | 1 |

1. Να γίνει πλήρης στατιστική ανάλυση, παρουσίαση και σχολιασμός των αποτελεσμάτων καθώς και έλεγχος των προϋποθέσεων (αναλυτική περιγραφική στατιστική, ANOVA, συγκρίσεις μέσων όρων, πίνακες, διαγράμματα κλπ).
2. Να κάνετε μια υπόθεση σχετικά με το τι συμβολίζουν, τι σημαίνουν, οι μεταβλητές D1, D2 και D3.
3. Να εκτιμήσετε την ευθεία παλινδρόμησης της απόδοσης Y πάνω στις D1 και D2. Να γίνει πλήρης στατιστική ανάλυση, παρουσίαση και σχολιασμός των αποτελεσμάτων καθώς και έλεγχος των προϋποθέσεων.
4. Να εκτιμήσετε την ευθεία παλινδρόμησης της απόδοσης Y πάνω στις D1 και D3. Να γίνει πλήρης στατιστική ανάλυση, παρουσίαση και σχολιασμός των αποτελεσμάτων καθώς και έλεγχος των προϋποθέσεων.
5. Να εκτιμήσετε την ευθεία παλινδρόμησης της απόδοσης Y πάνω στις D2 και D3. Να γίνει πλήρης στατιστική ανάλυση, παρουσίαση και σχολιασμός των αποτελεσμάτων καθώς και έλεγχος των προϋποθέσεων.
6. Με βάση τα αποτελέσματα από τις απαντήσεις στα ερωτήματα 3-5 σε τι συμπεράσματα/διαπιστώσεις καταλήγετε για τη σχέση μεταξύ της ANOVA και της Γραμμικής Παλινδρόμησης;

DO NOT COPY

## Άσκηση 12

Στο πείραμα της Άσκησης 11 μετρήθηκε και ένας δείκτης ποιότητας του παραγόμενου προϊόντος σε μία κλίμακα από 1 έως 10. Η τιμή 10 αντιστοιχεί στην πιο καλή ποιότητα. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 9.

**Πίνακας 9**

| | | Δείκτης | |
|---|---|---|---|
| **Λίπασμα** | **Απόδοση** | **Ποιότητας** | **Επανάληψη** |
| 1 | 750 | 5 | 1 |
| 1 | 740 | 7 | 2 |
| 1 | 755 | 6 | 3 |
| 1 | 745 | 8 | 4 |
| 1 | 730 | 9 | 5 |
| 2 | 690 | 8 | 1 |
| 2 | 650 | 5 | 2 |
| 2 | 630 | 4 | 3 |
| 2 | 610 | 3 | 4 |
| 2 | 660 | 7 | 5 |
| 3 | 450 | 5 | 1 |
| 3 | 490 | 6 | 2 |
| 3 | 500 | 7 | 3 |
| 3 | 520 | 8 | 4 |
| 3 | 430 | 4 | 5 |

1. Για τον δείκτη ποιότητας, να γίνει πλήρης στατιστική ανάλυση, παρουσίαση και σχολιασμός των αποτελεσμάτων καθώς και έλεγχος των προϋποθέσεων (αναλυτική περιγραφική στατιστική, ANOVA, συγκρίσεις μέσων όρων, πίνακες, διαγράμματα κλπ).
2. Να υπολογίσετε τον συντελεστή γραμμικής συσχέτισης (*r* του Pearson) μεταξύ της απόδοσης και του δείκτη ποιότητας.
3. Να αξιολογήσετε τη στατιστική σημαντικότητα και την ένταση/βαθμό του συντελεστή συσχέτισης.
   Υπόδειξη: Η στατιστική σημαντικότητα του συντελεστή συσχέτισης να αξιολογηθεί με παραμετρικό έλεγχο και με τις μεθόδους προσομοίωσης Monte-Carlo και Bootstrap.
4. Να εκτιμήσετε την ευθεία παλινδρόμησης του δείκτη ποιότητας πάνω στην απόδοση. Να γίνει πλήρης στατιστική ανάλυση, παρουσίαση και σχολιασμός των αποτελεσμάτων.
5. Για κάθε λίπασμα να απαντηθούν τα ερωτήματα 1 έως 3.
6. Να συγκρίνετε τις 3 ευθείες παλινδρόμησης (του δείκτη ποιότητας πάνω στην απόδοση, για κάθε λίπασμα).
   Υπόδειξη: Θα πρέπει να συγκριθούν οι σταθεροί όροι και οι κλίσεις των ευθειών.

DO NOT COPY

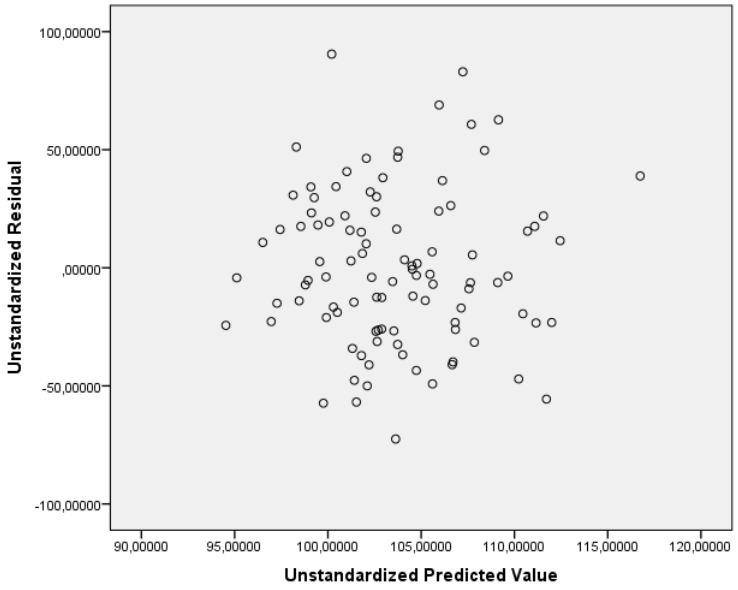
Να κάνετε το ίδιο και για τους 3 συντελεστές γραμμικής συσχέτισης (*r* του Pearson). Τι παρατηρείτε;

7. Να γίνει σύγκριση των 3 λιπασμάτων ως προς την ποιότητα αφού αφαιρεθεί η επίδραση της απόδοσης στην ποιότητα.
8. Να γίνει σύγκριση των 3 λιπασμάτων ως προς την ποιότητα θεωρώντας ότι η απόδοση ήταν σταθερή και ίση σε όλες τις μετρήσεις. Είναι ασφαλή τα αποτελέσματα της σύγκρισης;

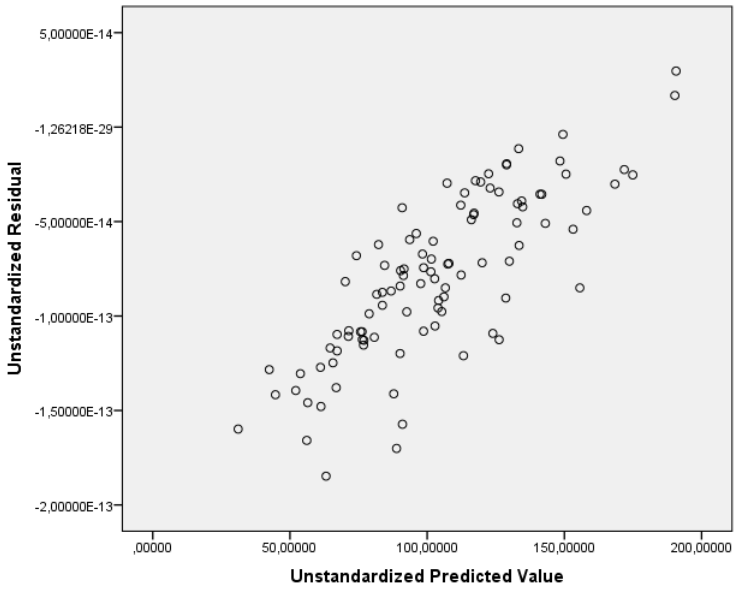
DO NOT COPY

## Άσκηση 13

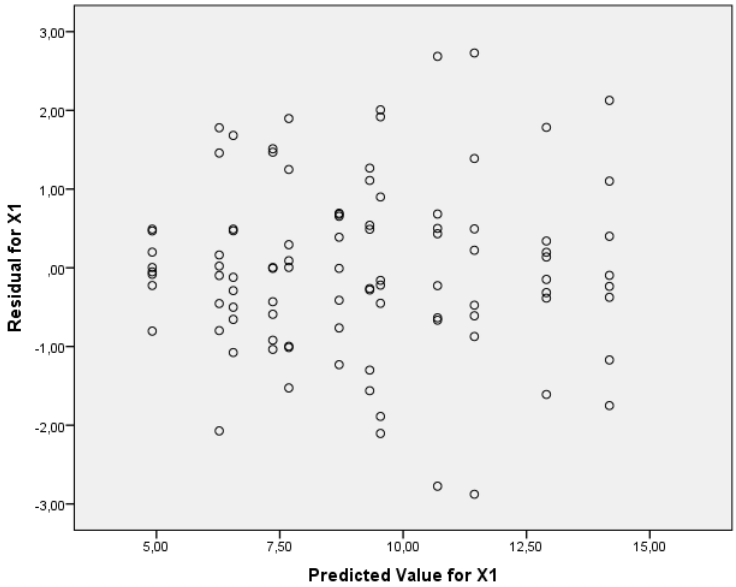
Κοιτάξτε με προσοχή τα παρακάτω διαγράμματα διασποράς (Unstandardized Predicted Values, Unstandardized Residuals) για τον έλεγχο της ομοσκεδαστικότητας των σφαλμάτων/υπολοίπων που προέκυψαν από μια σειρά γραμμικών υποδειγμάτων. Σε ποια από τα διαγράμματα φαίνεται να υπάρχει σοβαρή παραβίαση της ομοσκεδαστικότητας;



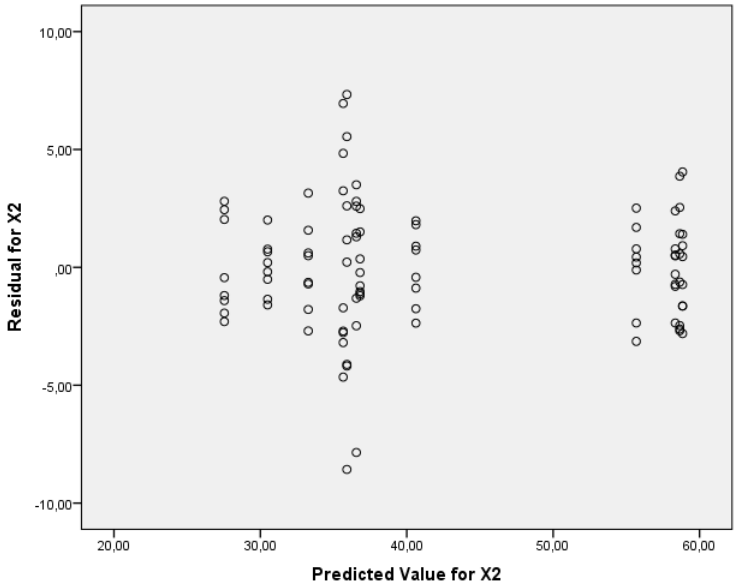

**Διάγραμμα 1**



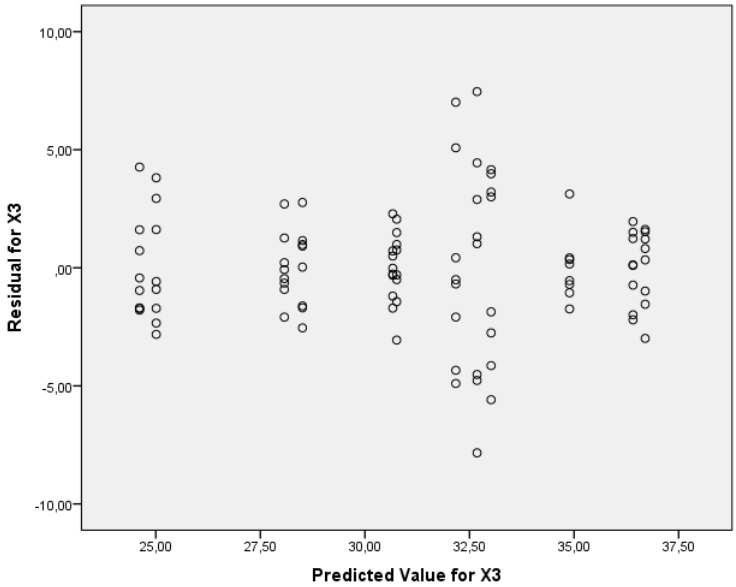

**Διάγραμμα 2**



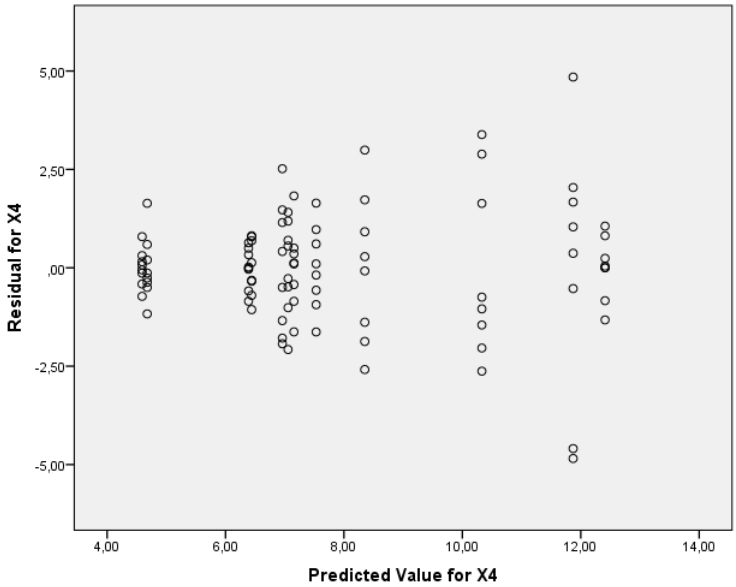

**Διάγραμμα 3**



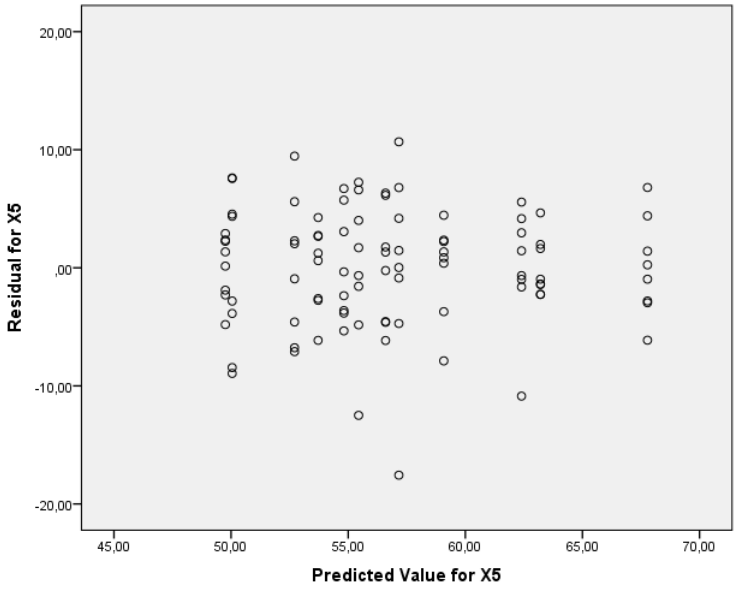

**Διάγραμμα 4**



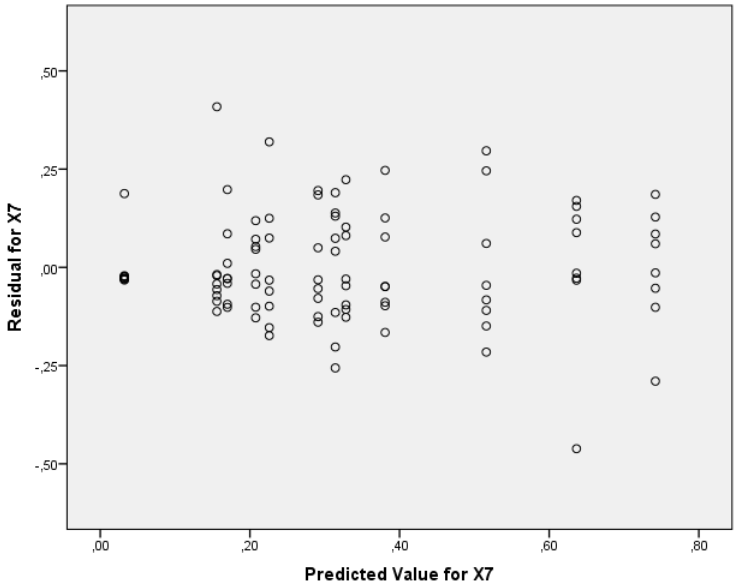

**Διάγραμμα 5**



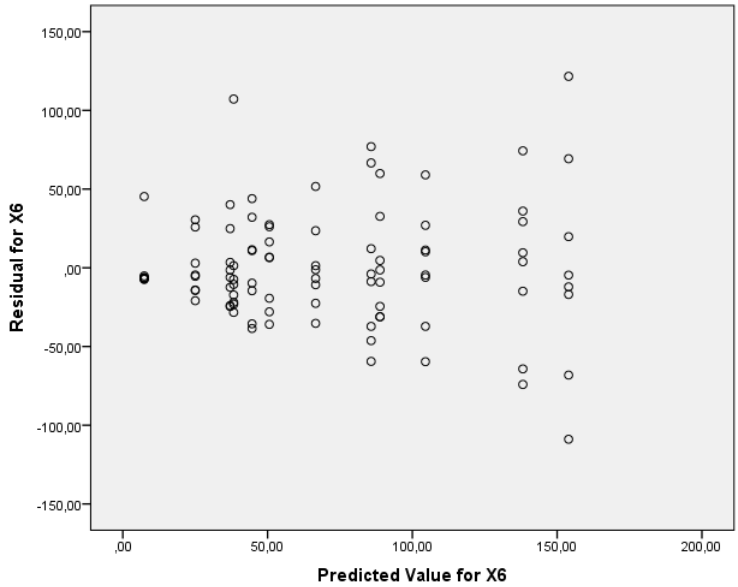

**Διάγραμμα 6**




**Διάγραμμα 7**



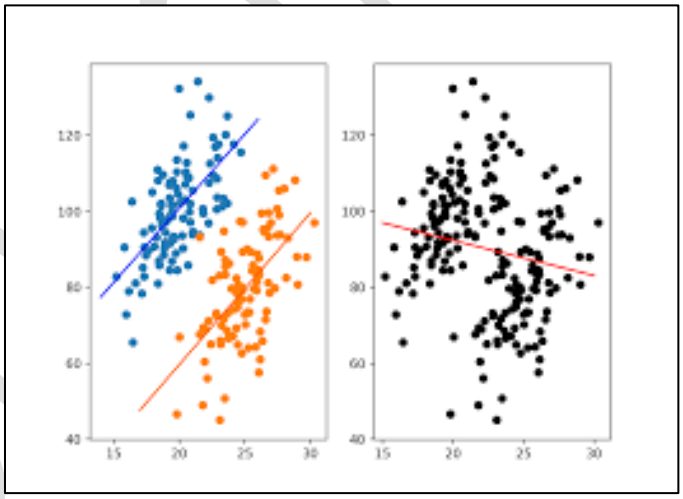

**Διάγραμμα 8**




**Διάγραμμα 9**

## Άσκηση 14

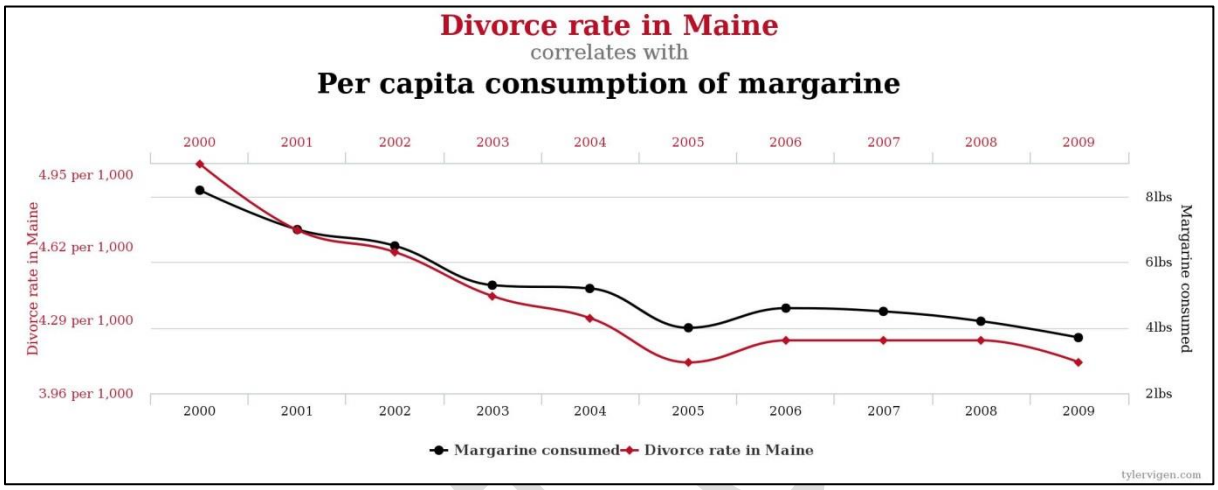
Μελετήστε με προσοχή το Διάγραμμα 10. Τι διαπιστώνετε; Πως ονομάζεται το "παράδοξο" που απεικονίζεται στο διάγραμμα;




**Διάγραμμα 10**

## Άσκηση 15

Μελετήστε με προσοχή το Διάγραμμα 11. Τι διαπιστώνετε; Μπορείτε να υποθέσετε κάποια αιτιώδη σχέση που να αιτιολογεί τη συμμεταβολή των δύο μεταβλητών (Divorce rate-Ρυθμός διαζυγίων και Consumption of margarine-κατανάλωση μαργαρίνης); Που μπορεί να οφείλεται αυτή η ισχυρή συμμεταβολή που είναι εντυπωσιακά εμφανής στο διάγραμμα;




**Διάγραμμα 11**

DO NOT