

**Βιοστατιστική
Παράδειγμα Υπολογισμού
Συντελεστού Συσχέτισης
και Ευθείας της Παλινδρόμησης**

Χαράλαμπος Γναρδέλλης

**Εφαρμογές Πληροφορικής στην Αλιεία και τις
Υδατοκαλλιεργείες**

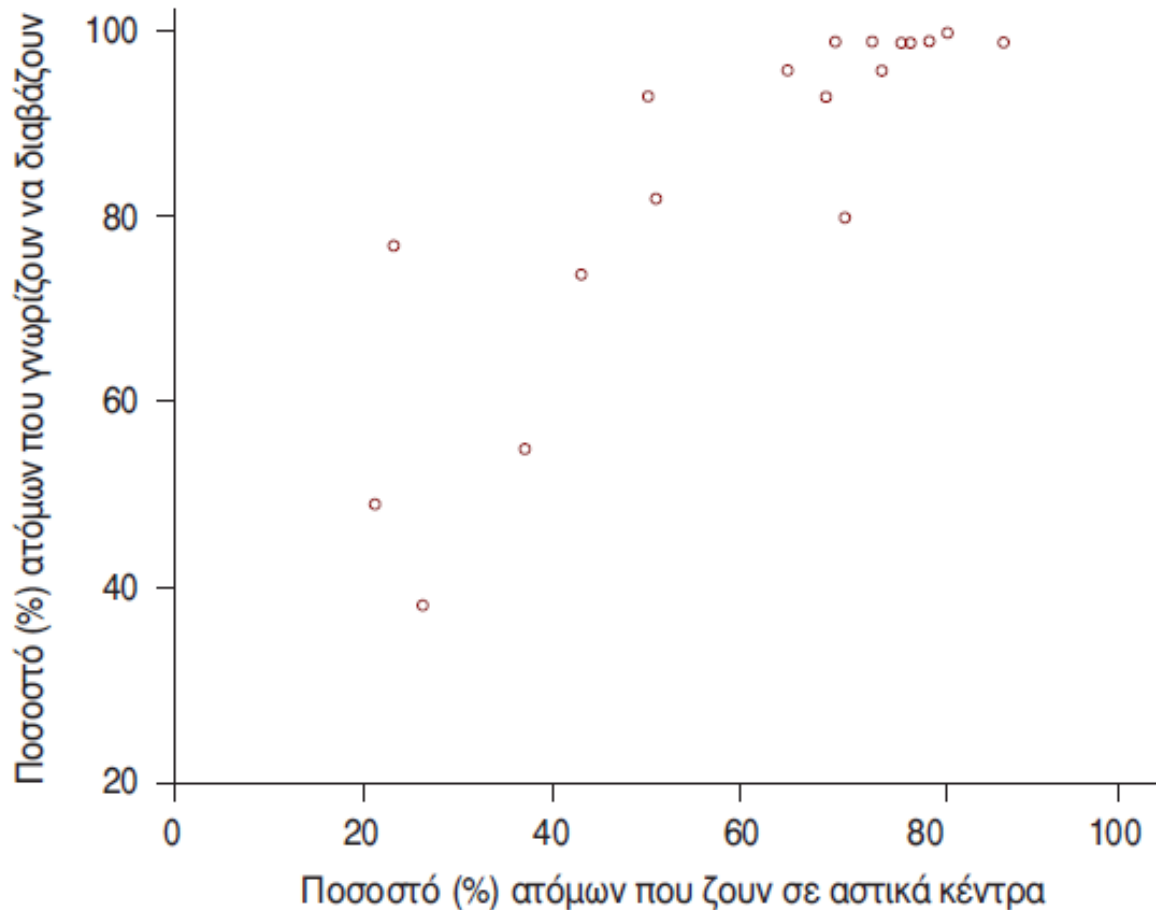
Υπολογισμός συντελεστού συσχέτισης

Ποσοστό ατόμων που ζουν σε αστικά κέντρα και ποσοστό ατόμων που γνωρίζουν να διαβάζουν σε ένα τυχαίο δείγμα χωρών του Οργανισμού Ηνωμένων Εθνών

	Ποσοστό (%) ατόμων που ζουν σε αστικά κέντρα	Ποσοστό (%) ατόμων που γνωρίζουν να διαβάζουν
Αίγυπτος	43	74
Αιθιοπία	21	49
Αφγανιστάν	26	38
Βολιβία	69	93
Γαλλία	80	99
Γερμανία	77	99
Γουατεμάλα	51	82
Δανία	88	99
Ελβετία	74	99
Ιράκ	71	80
Ιταλία	70	99
Καμπότζη	23	77
Νορβηγία	82	100
Πακιστάν	37	55
Πορτογαλία	65	96
Ταϊβάν	78	99
Ταϊλάνδη	50	93
Τουρκία	75	96

Κατασκευή διαγράμματος διασποράς

Διάγραμμα διασποράς του βαθμού αστικότητας και του ποσοστού των ατόμων που γνωρίζουν να διαβάζουν σε ένα τυχαίο δείγμα χωρών του Οργανισμού Ηνωμένων Εθνών



Συντελεστής Συσχέτισης

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]}}$$

Υπολογισμός μέσων τιμών

- για τη μεταβλητή X , που αναπαριστά το βαθμό αστικότητας των χωρών

$$\bar{x} = \frac{1}{18} \sum_{i=1}^{18} x_i = \frac{(43 + 21 + \dots + 75)}{18} = 60,0$$

- για τη μεταβλητή Y , που αναπαριστά το ποσοστό των ατόμων που γνωρίζουν να διαβάζουν

$$\bar{y} = \frac{1}{18} \sum_{i=1}^{18} x_i = \frac{(74 + 49 + \dots + 96)}{18} = 84,8$$

Υπολογισμός αθροίσματος γινομένων

- για τις δύο μεταβλητές X και Y

$$\begin{aligned} \sum_{i=1}^{18} (x_i - \bar{x})(y_i - \bar{y}) &= \\ &= (43 - 60)(74 - 84,8) + (21 - 60)(49 - 84,8) + \\ &+ \dots + (75 - 60)(96 - 84,8) = 6187,0 \end{aligned}$$

- και

$$\begin{aligned} \sum_{i=1}^{18} (x_i - \bar{x})^2 &= \\ &= (43 - 60)^2 + (21 - 60)^2 + \dots + (75 - 60)^2 = 8074,0, \\ \sum_{i=1}^{18} (y_i - \bar{y})^2 &= \\ &= (74 - 84,8)^2 + (49 - 84,8)^2 + \dots + (96 - 84,8)^2 = 6394,5. \end{aligned}$$

Υπολογισμός συντελεστού συσχέτισης

Άρα, ο συντελεστής συσχέτισης του Pearson ισούται με

$$r = \frac{\sum_{i=1}^{18} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{18} (x_i - \bar{x})^2 \right] \left[\sum_{i=1}^{18} (y_i - \bar{y})^2 \right]}} =$$
$$= \frac{6187}{\sqrt{(8074)(6394,5)}} = 0,86.$$

Υπολογισμός συντελεστών απλής παλινδρόμησης

Στο παράδειγμα του υπολογισμού του συντελεστού συσχέτισης μεταξύ βαθμού αστικότητας και βαθμού αλφαριθμητισμού των χωρών του ΟΗΕ θα κατασκευαστεί η ευθεία της παλινδρόμησης

$$\hat{y} = b_0 + b_1 x$$

με

εξαρτημένη μεταβλητή y τον αλφαριθμητισμό, και

ανεξάρτητη μεταβλητή x το βαθμό αστικότητας.

Οι συντελεστές του υποδείγματος υπολογίζονται από τις εξισώσεις

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Υπολογισμός των συντελεστών του υποδείγματος

Αρχικά υπολογίζουμε τις μέσες τιμές των δύο μεταβλητών

- για τη μεταβλητή X , που αναπαριστά το βαθμό αστικότητας των χωρών

$$\bar{x} = \frac{1}{18} \sum_{i=1}^{18} x_i = \frac{(43 + 21 + \dots + 75)}{18} = 60,0$$

- για τη μεταβλητή Y , που αναπαριστά το ποσοστό των ατόμων που γνωρίζουν να διαβάζουν

$$\bar{y} = \frac{1}{18} \sum_{i=1}^{18} x_i = \frac{(74 + 49 + \dots + 96)}{18} = 84,8$$

Στη συνέχεια υπολογίζουμε το άθροισμα των γινομένων των διαφορών για τις δύο μεταβλητές x , y και το άθροισμα των τετραγώνων των διαφορών για τη μεταβλητή x

- για τις δύο μεταβλητές X και Y

$$\begin{aligned} \sum_{i=1}^{18} (x_i - \bar{x})(y_i - \bar{y}) &= \\ &= (43 - 60)(74 - 84,8) + (21 - 60)(49 - 84,8) + \\ &+ \dots + (75 - 60)(96 - 84,8) = 6187,0 \end{aligned}$$

- και

$$\begin{aligned} \sum_{i=1}^{18} (x_i - \bar{x})^2 &= \\ &= (43 - 60)^2 + (21 - 60)^2 + \dots + (75 - 60)^2 = 8074,0 \end{aligned}$$

Αντικαθιστούμε τα προηγούμενα αποτελέσματα στις εξισώσεις

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Υπολογισμός του b_1

$$b_1 = \frac{\sum_{i=1}^{18} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{18} (x_i - \bar{x})^2} = \frac{6187}{8074} = 0,77$$

Υπολογισμός του b_0

$$b_0 = \bar{y} - b_1 \bar{x} = 84,8 - (0,77) \cdot (60) = 38,6$$

- Άρα η εξίσωση της παλινδρόμησης που εκτιμά το βαθμό αλφαριθμητισμού μιας χώρας με το βαθμό αστικότητας δίδεται από τη σχέση:

$$\hat{y} = 38,6 + 0,77x$$

Άσκηση

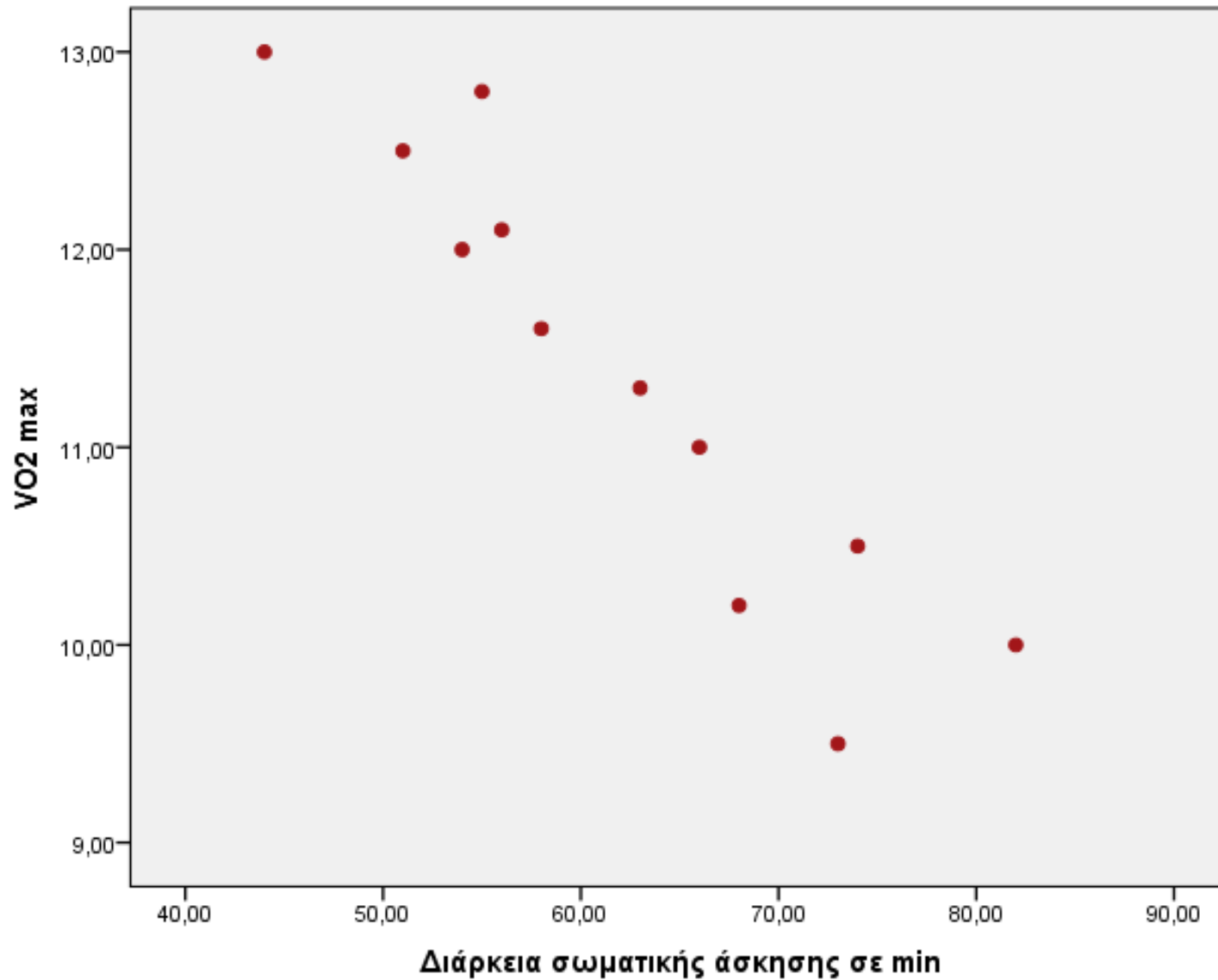
Στον παρακάτω πίνακα δίδονται οι τιμές του μέγιστου εισπνεόμενου όγκου οξυγόνου ($\text{VO}_2 \text{ max}$) 12 ενηλίκων έπειτα από εντατική σωματική άσκηση. Γενικά, ο μέγιστος εισπνεόμενος όγκος οξυγόνου αποτελεί μέτρο της εύρυθμης καρδιακής λειτουργίας και ελαττώνεται με την αύξηση του επιπέδου της σωματικής δραστηριότητας.

Άτομο	$\text{VO}_2 \text{ max}$	Διάρκεια σωματικής άσκησης σε min
1	82	10,0
2	73	9,5
3	68	10,2
4	74	10,5
5	66	11,0
6	63	11,3
7	58	11,6
8	54	12,0
9	56	12,1
10	51	12,5
11	55	12,8
12	44	13,0

Να γίνει το διάγραμμα σκέδασης του μέγιστου εισπνεόμενου όγκου οξυγόνου με τη διάρκεια της σωματικής άσκησης.

Να προσδιοριστεί η ευθεία των ελαχίστων τετραγώνων για τη σχέση του $\text{VO}_2 \text{ max}$ με τη διάρκεια της σωματικής άσκησης.

Διάγραμμα διασποράς των δύο μεταβλητών



Υπολογισμός των παραμέτρων του υποδείγματος

n	$\text{VO}_2 \text{ max}$ x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	Διάρκεια σωματικής άσκησης σε $\text{min } y_i$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
1	82	20	400	10,0	-1,38	-27,5
2	73	11	121	9,5	-1,88	-20,625
3	68	6	36	10,2	-1,18	-7,05
4	74	12	144	10,5	-0,88	-10,5
5	66	4	16	11,0	-0,38	-1,5
6	63	1	1	11,3	-0,07	-0,075
7	58	-4	16	11,6	0,23	-0,9
8	54	-8	64	12,0	0,63	-5
9	56	-6	36	12,1	0,73	-4,35
10	51	-11	121	12,5	1,13	-12,375
11	55	-7	49	12,8	1,43	-9,975
12	44	-18	324	13,0	1,63	-29,25
	$\bar{x} = 62$		$\sum_{i=1}^{12} (x_i - \bar{x})^2$ =1328		$\bar{y} = 11,38$	$\sum_{i=1}^{12} (x_i - \bar{x})(y_i - \bar{y})$ =-129,1

$$b_1 = \frac{\sum_{i=1}^{12} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{12} (x_i - \bar{x})^2} = \frac{-129,1}{1328} = -0,009$$

$$b_0 = \bar{y} - b_1 \bar{x} = 11,38 - (-0,09)(62) = 16,96$$

άρα $\hat{y} = b_0 + b_1 x = 16,96 - 0,09x$

Στον παρακάτω πίνακα δίδονται οι τιμές μεθυλικού υδραργύρου από διατροφική πρόσληψη και οι αντίστοιχες συγκεντρώσεις υδραργύρου στο αίμα 12 ατόμων που εκτέθηκαν σε μεθυλικό υδράργυρο λόγω κατανάλωσης μολυσμένων ψαριών:

Μεθυλικός υδράργυρος
από πρόσληψη (μg Hg/ημέρα)

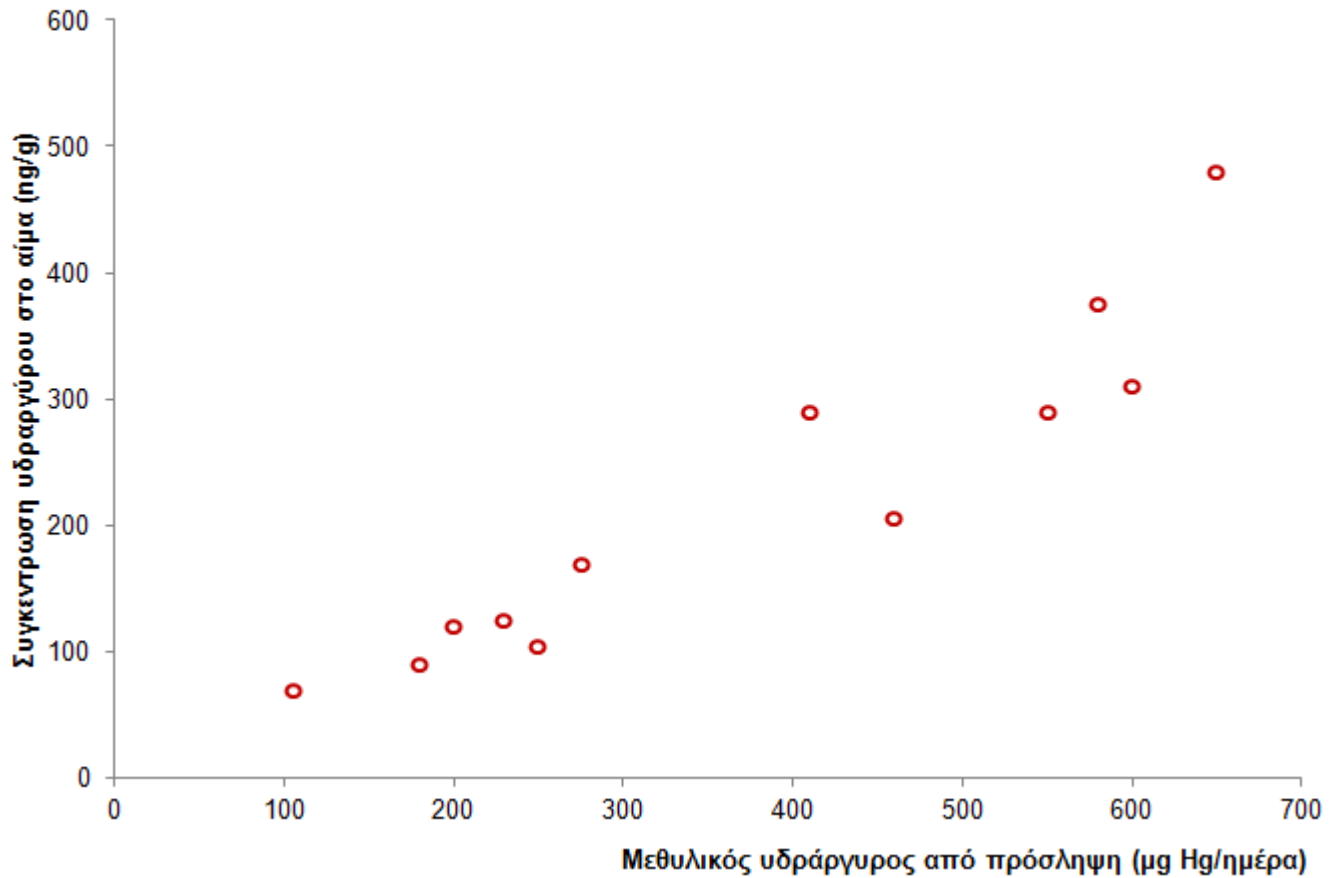
Συγκέντρωση υδραργύρου
στο αίμα (ng/g)

180	90
200	120
230	125
410	290
600	310
550	290
275	170
580	375
105	70
250	105
460	205
650	480

Να κατασκευαστεί η εξίσωση της παλινδρόμησης με εξαρτημένη μεταβλητή τη συγκέντρωση υδραργύρου στο αίμα και ανεξάρτητη τον μεθυλικό υδράργυρο από τη διατροφική πρόσληψη .

Κατασκευή διαγράμματος διασποράς

Διάγραμμα διασποράς



Οι συντελεστές του υποδείγματος $\hat{y} = b_0 + b_1 x$

υπολογίζονται από τις εξισώσεις

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Όπου y η συγκέντρωση υδραργύρου στο αίμα και
 x η πρόσληψη μεθυλικού υδραργύρου από τη διατροφή

Υπολογισμός των συντελεστών του υποδείγματος

Υπολογισμός μέσων τιμών για τις δύο μεταβλητές

Μεθυλικός υδράργυρος

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{180 + 200 + 230 + \dots + 460 + 650}{12} = 374,2$$

Συγκέντρωση υδραργύρου στο αίμα

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{90 + 120 + 125 + \dots + 205 + 480}{12} = 219,2$$

Υπολογισμός αθροίσματος γινομένων

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$= (180 - 374,2)(90 - 219,2) + (200 - 374,2)(120 - 219,2) + \dots + (650 - 374,2)(480 - 219,2)$$

$$= (-194,2)(-129,2) + (-174,2)(-99,2) + \dots + (275,8)(260,8)$$

$$= 25090,6 + 17280,6 + \dots + 71928,6$$

$$= 253441,7$$

Υπολογισμός των τετραγώνων των διαφορών του x

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})^2 &= (180 - 374,2)^2 + (200 - 374,2)^2 + \dots + (460 - 374,2)^2 + (65 - 374,2)^2 \\ &= 395541,7\end{aligned}$$

Αντικαθιστούμε τα προηγούμενα αποτελέσματα στις εξισώσεις

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Υπολογισμός του b_1

$$b_1 = \frac{\sum_{i=1}^{12} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{12} (x_i - \bar{x})^2} = \frac{253441,7}{395541,7} = 0,64$$

Υπολογισμός του b_0

$$b_0 = \bar{y} - b_1 \bar{x} = 219,2 - (0,64) \cdot (374,2) = 219,2 - 239,49 = -20,29$$

Άρα η εξίσωση της παλινδρόμησης που εκτιμά τη συγκέντρωση του υδραργύρου στο αίμα σε σχέση με την πρόσληψη μεθυλικού υδραργύρου από τη διατροφή, δίδεται από τη σχέση:

$$\hat{y} = -20,29 + 0,64x$$