

Βιοστατιστική

Παράδειγμα Υπολογισμού Συντελεστού Συσχέτισης

Χαράλαμπος Γναρδέλλης

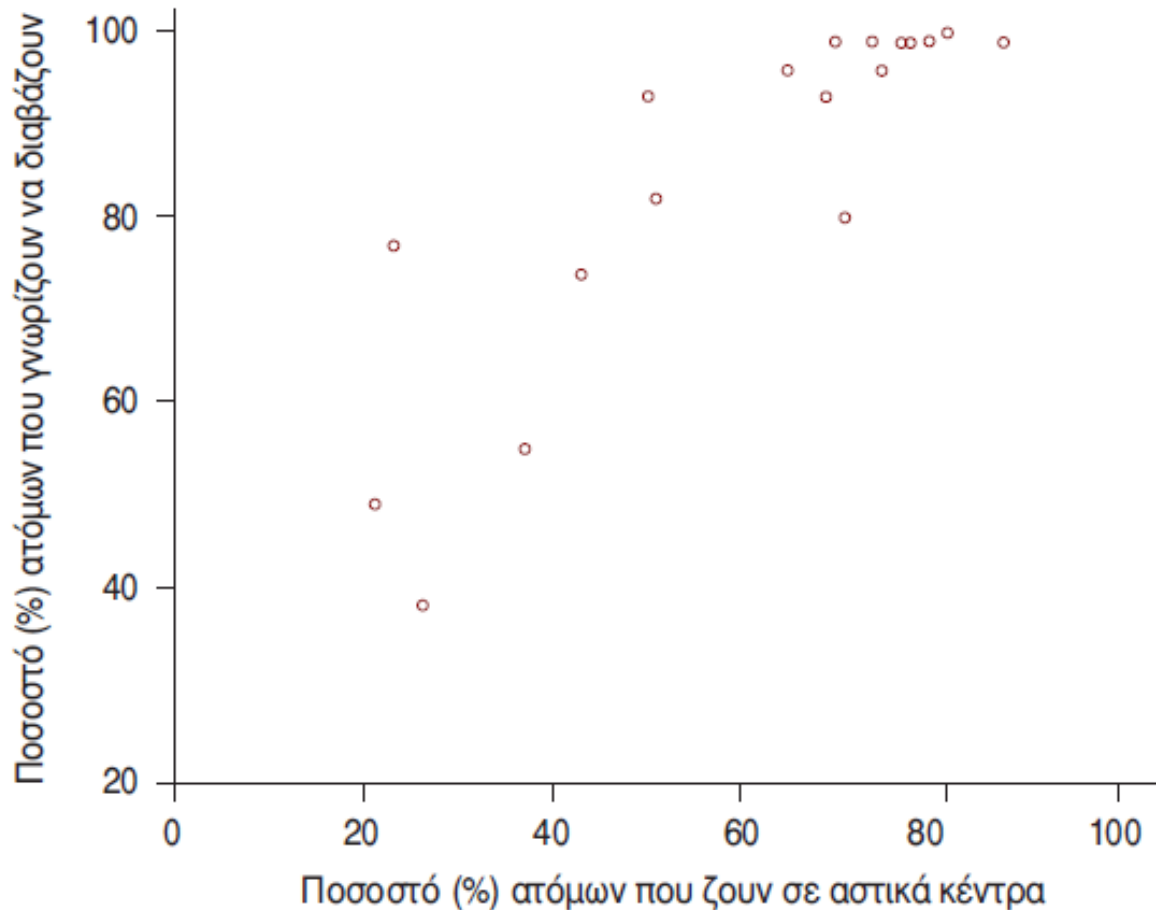
Υπολογισμός συντελεστού συσχέτισης

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

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

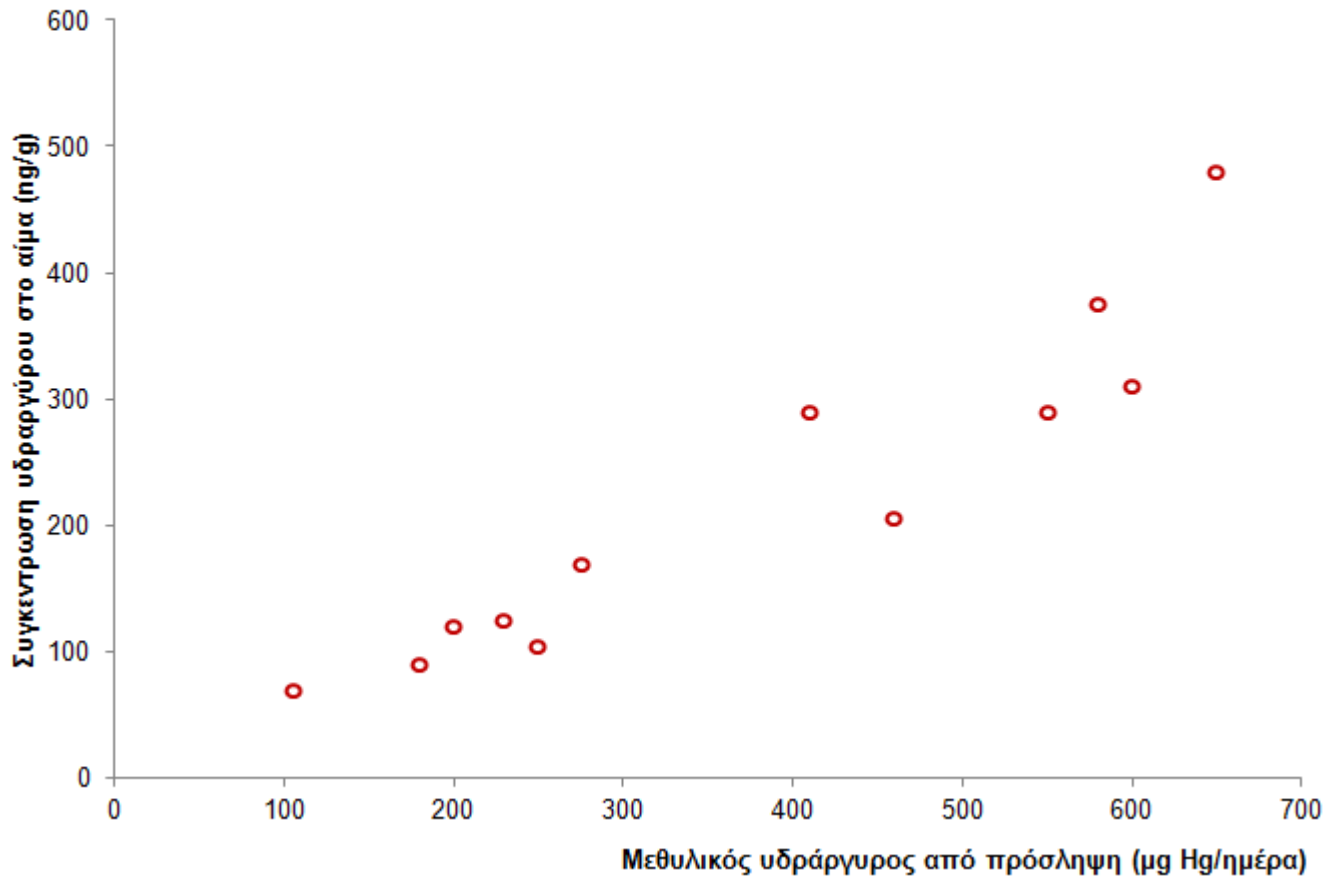
Στον παρακάτω πίνακα δίδονται οι τιμές μεθυλικού υδραργύρου από διατροφική πρόσληψη και οι αντίστοιχες συγκεντρώσεις υδραργύρου στο αίμα 12 ατόμων που εκτέθηκαν σε μεθυλικό υδράργυρο λόγω κατανάλωσης μολυσμένων ψαριών:

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 διερευνηθεί αν υπάρχει σχέση μεταξύ των δύο μεταβλητών με τη βοήθεια του συντελεστή συσχέτισης του Pearson. Να ορισθεί το είδος της συσχέτισης.

Κατασκευή διαγράμματος διασποράς

Διάγραμμα διασποράς



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

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

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 τυπικών αποκλίσεων

$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{(180 - 314,2)^2 + (200 - 314,2)^2 + \dots + (460 - 314,2)^2 + (65 - 314,2)^2}{12-1}}$$

$$= \sqrt{\frac{395541,7}{11}} = \sqrt{35958,3} = 189,6$$

$$s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}$$

$$= \sqrt{\frac{(90 - 219,2)^2 + (120 - 219,2)^2 + \dots + (205 - 219,2)^2 + (480 - 219,2)^2}{12-1}}$$

$$= \sqrt{\frac{183891,7}{11}} = \sqrt{16717,4} = 129,3$$

Υπολογισμός συντελεστού συσχέτισης

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{253441,7}{11 \cdot (189,6) \cdot (129,3)} = 0,94$$

Άρα η σχέση των δύο μεταβλητών είναι έντονα θετική.

Άσκηση

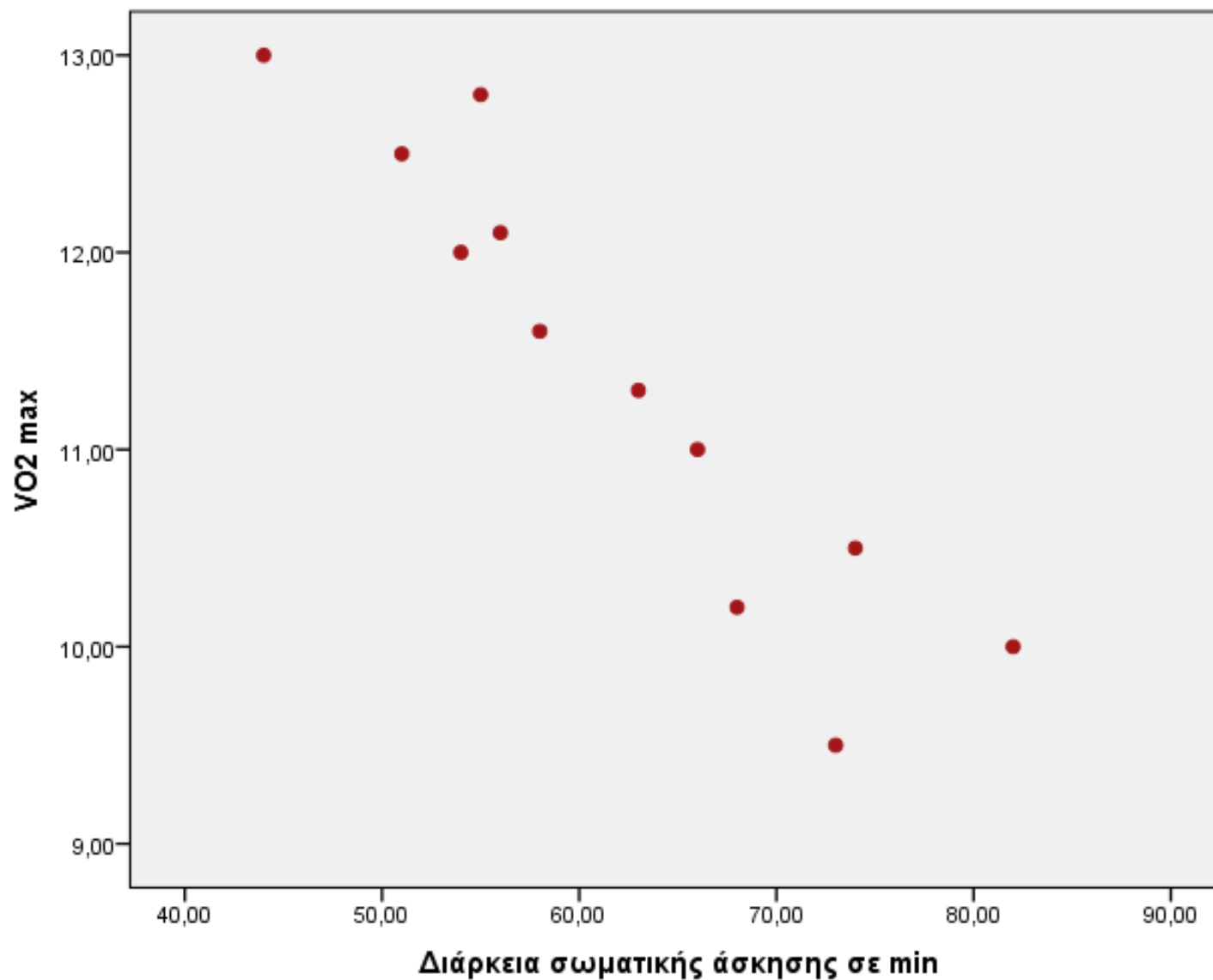
Στον παρακάτω πίνακα δίδονται οι τιμές του μέγιστου εισπνεόμενου όγκου οξυγόνου ($\text{VO}_2 \text{ max}$) 12 ενηλίκων έπειτα από εντατική σωματική άσκηση. Γενικά, ο μέγιστος εισπνεόμενος όγκος οξυγόνου αποτελεί μέτρο της εύρυθμης καρδιακής λειτουργίας και ελαττώνεται με την αύξηση του επιπέδου της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

Διάγραμμα διασποράς των δύο μεταβλητών



Υπολογισμός μέσων τιμών για τις δύο μεταβλητές

Εισπνεόμενος όγκος οξυγόνου $\text{VO}_2 \text{ max}$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{82 + 73 + 68 + \dots 55 + 44}{12} = 62$$

Διάρκεια σωματικής άσκησης

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{10 + 9,5 + 10,2 + \dots 12,8 + 13}{12} = 11,38$$

Υπολογισμός αθροίσματος γινομένων

$$\begin{aligned} & \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= (83 - 62)(10 - 11,38) + (73 - 62)(9,5 - 11,38) + \dots + (44 - 62)(13 - 11,38) \\ &= (20)(-1,38) + (11)(-1,88) + \dots + (-18)(1,63) \\ &= -27,6 + (-20,68) + \dots + (-29,34) \\ &= -129,1 \end{aligned}$$

Υπολογισμός τυπικών αποκλίσεων

$$\begin{aligned} s_x &= \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \\ &= \sqrt{\frac{(82 - 62)^2 + (730 - 62)^2 + \dots + (55 - 62)^2 + (44 - 62)^2}{12 - 1}} \\ &= \sqrt{\frac{1328}{11}} = \sqrt{120,73} = 10,99 \end{aligned}$$

$$\begin{aligned} s_y &= \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} \\ &= \sqrt{\frac{(10 - 11,38)^2 + (9,5 - 11,38)^2 + \dots + (12,8 - 11,38)^2 + (13 - 11,38)^2}{12 - 1}} \\ &= \sqrt{\frac{14,72}{11}} = \sqrt{1,338} = 1,156 \end{aligned}$$

Υπολογισμός συντελεστού συσχέτισης

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{-129,1}{11 \cdot (10,99) \cdot (1,156)} = -0,92$$

Άρα η σχέση των δύο μεταβλητών είναι έντονα αρνητική.