

Άσκηση 14

Στον πίνακα που ακολουθεί δίνονται οι ώρες μελέτης 10 φοιτητών για ένα ΤΕΓΤ και η επίδοσή τους στο ΤΕΓΤ.

Ωρες Μελέτης (x_i)	7	8	4	9	13	5	9	6	16	3	Σύνολο
Επίδοση (y_i)	70	76	57	77	91	66	82	64	96	50	80
$x_i y_i$	490	608	228	693	1183	330	738	384	1536	150	729
x_i^2	49	64	16	81	169	25	81	36	256	9	6340
y_i^2	4.900	5776	3249	5929	8281	4356	6724	4096	9216	2500	786
											55027

Να υπολογίσετε τον συντελεστή συσχέτισης Pearson και να χαρακτηρίσετε το είδος της συσχέτισης.

$$r = \frac{6340 - 10 \cdot 8 \cdot 72,9}{\sqrt{786 - 10 \cdot 8^2} \sqrt{55027 - 10 \cdot 72,9^2}} \Rightarrow$$

$$r = \frac{6340 - 5838}{\sqrt{786 - 640} \sqrt{55027 - 53144,1}} \Rightarrow$$

$$r = \frac{508}{\sqrt{46} \sqrt{1883,9}} \Rightarrow$$

$$r = \frac{508}{12,08 \cdot 43,39} \Rightarrow$$

$$r = \frac{508}{524,18} \Rightarrow r \approx 0,96$$

Πολύ ισχυρή γραμμική συσχέτιση.

Άσκηση 15

Δίνονται οι ώρες εργασίας 10 φοιτητών της εδοκμάδα πριν από το test και
 επίδοσή τους. Όπου:

x_i : ώρες εργασίας

y_i : επίδοσή

Ωρες Εργασίας (x_i)	20	15	4	30	20	15	8	30	4	0
Επίδοσή (y_i)	87	71	89	60	77	89	91	65	94	92

Να υπολογιστεί ο συντελεστής Pearson (r).

Λύση:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{146}{10} = 14,6$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{815}{10} = 81,5$$

$x_i y_i$	1740	1065	356	1800	1540	1335	728	1950	376	0	Σύνολο
x_i^2	400	225	16	900	400	225	64	900	16	0	10.830
y_i^2	7569	5041	7921	3600	5929	7921	8281	4995	8836	0	67.787

$$r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum x_i^2 - n \cdot \bar{x}^2} \sqrt{\sum y_i^2 - n \cdot \bar{y}^2}} \Rightarrow r = \frac{10.830 - 10 \cdot 14,6 \cdot 81,5}{\sqrt{3.146 - 10 \cdot 14,6^2} \sqrt{67.787 - 10 \cdot 81,5^2}} \Rightarrow$$

$$r = \frac{10.830 - 11.899}{\sqrt{3.146 - 2131,6} \sqrt{67.786 - 66.422,5}} \Rightarrow r = \frac{-1069}{\sqrt{1014,4} \sqrt{1.364,5}} \Rightarrow$$

$$r = \frac{-1069}{31,84 \cdot 36,96} \Rightarrow r = \frac{-1069}{1175,85} \Rightarrow r \approx -0,91$$

Υπάρχει πολύ ισχυρή αρνητική

Άσκηση 16

Επιλέξαμε ένα δείγμα μεγέθους $n=10$ και εξετάζουμε ως προς δύο μεταβλητές X και Y .

											Σύνολο
x_i	1	2	2	3	3	3	4	4	5	5	32
y_i	4	3	4	2	5	5	6	7	6	8	50

$$\text{Δίνονται: } \sum x_i^2 = 118, \sum x_i y_i = 176, \sum y_i^2 = 280$$

Να βρείτε τον συντελεστή συσχέτισης και να ερμηνεύσετε την συσχέτιση.

Λύση:

$$\bar{x} = \frac{32}{10} = 3,2$$

$$\bar{y} = \frac{50}{10} = 5$$

											Σύνολο
$x_i y_i$	4	6	8	6	15	24	28	30	40		176
x_i^2	1	4	4	9	9	16	16	25	25		118
y_i^2	16	9	16	4	25	36	49	36	64		280

$$r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum x_i^2 - n \cdot \bar{x}^2} \sqrt{\sum y_i^2 - n \cdot \bar{y}^2}} \Rightarrow r = \frac{176 - 10 \cdot 3,2 \cdot 5}{\sqrt{118 - 10 \cdot 3,2^2} \sqrt{280 - 10 \cdot 5^2}} =$$

$$r = \frac{176 - 160}{\sqrt{118 - 102,4} \sqrt{280 - 250}} \Rightarrow r = \frac{16}{\sqrt{15,6} \sqrt{30}} \Rightarrow$$

$$r = \frac{16}{3,94 \cdot 5,47} \Rightarrow r = \frac{16}{21,55} \Rightarrow r = 0,74$$

Άρα, έχουμε ισχυρή συσχέτιση

Άσκηση 17

Στον πίνακα που ακολουθεί δίνεται ο βαθμός στατιστικής σε μια πρόδο και βαθμός στις τελικές εξετάσεις σε ένα δείγμα 20 φοιτητών.

Φοιτητής	Βαθμός Πρόδοι (x _i)	Βαθμός Εξετάσεων (y _i)
1	65	85
2	50	77
3	75	90
4	70	84
5	68	61
6	58	70
7	49	76
8	92	78
9	68	80
10	93	92
11	38	56
12	92	95
13	59	60
14	66	87
15	61	83
16	71	77
17	68	79
18	82	74
19	85	94
20	67	80

α) Να κατασκευάσετε διάγραμμα διασποράς

β) Να υπολογίσετε τον συντελεστή Pearson

γ) Να βρείτε την ευθεία ελαχίστων τετραγώνων του δείγματος.

δ) Να προβλέψετε την μέση βαθμολογία στις τελικές εξετάσεις για κάποιον που στην πρόοδο έγραψε 40.

Λύση:

α) Το διαγράμμά θα γίνει μέσω excel. $\rightarrow$ μαζεύουμε τις τιμές $x_i, y_i \rightarrow$ έργο $\rightarrow$ Διασπορά.

$$\theta) r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum x_i^2 - n \cdot \bar{x}^2} \sqrt{\sum y_i^2 - n \cdot \bar{y}^2}} \Rightarrow r = \frac{n \cdot \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \cdot \sum x_i^2 - (\sum x_i)^2} \sqrt{n \cdot \sum y_i^2 - (\sum y_i)^2}} \Rightarrow$$

$$r = \frac{20 \cdot 110579 - 1377 \cdot 1578}{\sqrt{20 \cdot 98945 - (1377)^2} \sqrt{20 \cdot 126756 - (1578)^2}} \Rightarrow r = 0,63343$$

Έχουμε μέση βαθμολογία.

$$\gamma) \hat{b}_1 = \frac{n \cdot \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \cdot \sum x_i^2 - (\sum x_i)^2} \Rightarrow \hat{b}_1 = \frac{20 \cdot 110579 - (1377 \cdot 1578)}{20 \cdot 98945 - (1377)^2} \Rightarrow$$

$\downarrow$ με στρογγυλοποίηση.

$$\hat{b}_1 = 0,46724 \approx 0,467$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} \Rightarrow \hat{b}_0 = \frac{\sum y_i}{n} - \hat{b}_1 \frac{\sum x_i}{n}$$

$$\Rightarrow \hat{b}_0 = \frac{1578}{20} - 0,47 \frac{1377}{20}$$

$$\Rightarrow \hat{b}_0 = 78,9 - 0,47 \cdot 68,85$$

$$\Rightarrow \hat{b}_0 = 78,9 - 32,35 \Rightarrow \hat{b}_0 = 46,55$$

Άρα, $\hat{y} = 46,55 + 0,47x$.

δ) Αν $x=40$, τότε:

$$\hat{y} = 46,55 + 0,47 \cdot 40 \Rightarrow$$

$$\hat{y} = 46,55 + 18,8 \Rightarrow$$

$$\hat{y} = 65,35$$

Δίνουμε η πρόβλεψη του Μ.Ο. των λαθών για όσους έγραψαν 40 στην πρόδο.
Αλλ, όσοι έγραψαν 40 περιμέναμε ότι κατά Μ.Ο. στις εξετάσεις θα γράφουν 65,35.

Άσκηση 18

Θέλουμε να εξετάσουμε αν συσχετίζονται οι τυχαίες μεταβλητές x, y όπου:

x_i : ο αριθμός παραποδοθέντων τηλεόρασης υπερμέγας από μαθητές

y_i : ο αριθμός μελέτης ωρών υπερμέγας.

Για να το πετύχαμε επιλέξαμε ένα τυχαίο δείγμα 10 μαθητών.

x_i	0	0	1	2	1	6	4	2	3	1
y_i	5	7	4	4	3	0	1	3	2	4

$$\text{Δίνονται: } \sum x_i y_i = 35, \sum x_i^2 = 72, \sum y_i^2 = 145$$

i) Να υπολογιστεί και να ερμηνευθεί ο συντελεστής συσχέτισης.

ii) Να βρεθεί η ευθεία γραμμικής παλινδρόμησης ελαττώσεων τετραγώνων της πάνω στην x .

iii) Να εκτιμήσετε από την ευθεία γραμμικής παλινδρόμησης που βρήκατε, τις μελέτες ενός μαθητή αν σας πει ότι παραποδοθεί 5 ώρες τηλεόραση υπερμέγας.

Λύση

$$a) \sum x_i = 20 \text{ και } \bar{x} = \frac{20}{10} = 2$$

$$\sum y_i = 33 \text{ και } \bar{y} = \frac{33}{10} = 3,3$$

$$r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum x_i^2 - n \cdot \bar{x}^2} \sqrt{\sum y_i^2 - n \cdot \bar{y}^2}} \Rightarrow r = \frac{35 - 10 \cdot 2 \cdot 3,3}{\sqrt{72 - 10 \cdot 2^2} \sqrt{143 - 10 \cdot 3,3^2}} \Rightarrow$$

$$r = \frac{35 - 66}{\sqrt{72 - 40} \sqrt{143 - 108,9}} \Rightarrow r = \frac{-31}{\sqrt{32} \sqrt{36,1}} \Rightarrow r = \frac{-31}{3,66 \cdot 6} \Rightarrow r = -0,91$$

πολύ ισχυρή
(αρνητική)
συσχέτιση

β)

$$\hat{b}_1 = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum x_i^2 - n \cdot \bar{x}^2} = \frac{-31}{32} = -0,9687$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} \Rightarrow \hat{b}_0 = 3,3 + 0,9687 \cdot 2 \Rightarrow \hat{b}_0 = 3,3 + 1,9374$$

$$\Rightarrow \hat{b}_0 = 5,2374$$

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x \Rightarrow \hat{y} = 5,2374 - 0,9687 x$$

$$\gamma) \hat{y} = 5,2374 - 0,9687 \cdot 5 \Rightarrow$$

$$\hat{y} = 5,2374 - 4,8433 \Rightarrow$$

$$\hat{y} = 0,3939$$

Άσκηση (13)

Ένας γεωργός που καλλιεργεί σιτάρι συγκεντρώνει τα παρακάτω δεδομένα,
οποία δίνουν την απόδοση ανά στρέμμα (kg) που προέκυψε από την χρήση x
φορτικών ποσοτήτων λιπάσματος (kg/στρέμμα).

Λιπάσμα (x_i)	0	40	80	100	150	180	200	250	Σύνολο
Απόδοση (y_i)	60	70	100	130	170	180	220	230	1160

- Να εξετάσετε αν υπάρχει γραμμική συσχέτιση ως 2 μεταβλητές.
- Να εκτιμήσετε την ευθεία παλινδρόμησης x και y .
- Να προβλέψετε την μέση απόδοση των καλλιεργιών αν ρίξετε:
 - 180 kg λιπάσμα/στρέμμα
 - 400 kg λιπάσμα/στρέμμα

Λύση:

$$a) r = \dots = 0,985 \text{ υπάρχει πολύ ισχυρή}$$

$$b) b_1 = \dots = 0,757$$

$$b_0 = \dots = 50,257$$

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

$$\hat{y} = 50,257 + 0,757x$$

$$g) i) \text{ για } x = 180 \text{ kg}$$

$$\hat{y} = 50,257 + 0,757 \cdot 180 \Rightarrow \hat{y} = 50,257 + 136,26 \Rightarrow \hat{y} = 186,517$$

$$E(Y|X=120) = 141,097$$

ii) Επειδή οι τιμές του δείγματος είναι από 0 έως 250 κιλά, δεν μπορεί να γίνει ασφαλή πρόβλεψη για $x=400$ κιλά.

Άσκηση 20

Στον παρακάτω πίνακα δίνεται το ποσό (X) των χρεώσεων (σε χιλιάδες €) που ζήτησε μια εταιρία για διαφήμιση τους 10 προηγούμενους μήνες και αντίστοιχα κέρδη της (Y) σε χιλιάδες €.

											Σύνολο
Έξοδα Διαφήμισης (X_i)	2	1	2	3	1	4	3	4	5	5	30
Κέρδη (Y_i)	10	6	11	12	5	15	10	11	16	14	110

Δίνεται: $\sum Y_i^2 = 1324$

- Να βρεθεί ο συντελεστής Pearson (r)
- Να εκτιμηθεί η ευθεία παλινδρόμησης των κερδών πάνω στα έξοδα διαφήμισης.
- Να βρείτε τη ποσοστό μεταβλητότητας των κερδών εξηγείται από τη γραμμική σχέση της με τα έξοδα (υπολογίζοντας το r^2 και μετατρέποντας το σε ποσοστό)
- Να δώσετε ένα 90% διάστημα εμπιστοσύνης.
α) για το μέσο μηνιαίο κέρδος των εταιριών που ζοδεύουν 2.500 €/μήνα για διαφήμιση.
β) για το μηνιαίο κέρδος μιας εταιρίας από τον επόμενο μήνα αναμένεται να ζοδεύει 2.500 € για διαφήμιση.

Λύση:

i)

Σύνολο

$x_i y_i$	20	6	22	36	5	60	30	44	80	70	373
x_i^2	4	1	4	9	1	16	9	16	25	25	110

$$\bar{x} = \frac{30}{10} = 3$$

$$\bar{y} = \frac{110}{10} = 11$$

$$r = \frac{\sum x_i y_i - v \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum x_i^2 - v \cdot \bar{x}^2} \sqrt{\sum y_i^2 - v \cdot \bar{y}^2}} \Rightarrow r = \frac{373 - 10 \cdot 3 \cdot 11}{\sqrt{110 - 10 \cdot 3^2} \sqrt{1324 - 10 \cdot 11^2}} \Rightarrow$$

$$r = \frac{373 - 330}{\sqrt{110 - 90} \sqrt{1324 - 1210}} \Rightarrow r = \frac{43}{\sqrt{20} \sqrt{114}} \Rightarrow$$

$$r = \frac{43}{4,47 \cdot 10,67} \Rightarrow r = \frac{43}{47,7396} \Rightarrow r \approx 0,90$$

Άρα, υπάρχει πολύ ισχυρή γραμμική συσχέτιση.

ii)

$$\hat{b}_1 = \frac{\sum x_i y_i - v \cdot \bar{x} \cdot \bar{y}}{\sum x_i^2 - v \cdot \bar{x}^2} = \frac{43}{20} = 2,15$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 11 - 2,15 \cdot 3 = 11 - 6,45 = 4,55$$

Άρα, η ευθεία ελαχίστων τετραγώνων της Y στην X είναι:

$$\hat{Y} = 4,55 + 2,15X$$

iii) $r^2 = 0,9^2 = 0,81$ ή 81%.

Το 81% των κερδών οφείλεται στην γραμμική τμή εξάρτηση από τα έξοδα για διαφήμιση. ~~Το 81% της απόδοσης που θα πάρει και το υπόλοιπο 19% είναι το εφάρμο~~

iv) Για $x = 2,5$ χιλ. € έχουμε

$$\hat{y}_0 = 4,55 + 2,21 \cdot 2,5$$

$$\hat{y}_0 = 4,55 + 5,525$$

$$\hat{y}_0 = 9,925 \text{ χιλ. € ή } 9.925 \text{ €}$$

Είναι η συστηματική εκτίμηση των κερδών, αν τα έξοδα διαφήμισης είναι 2,5 χιλ. €. Αυτά, από όλες αυτές τις εταιρίες η μέση αλήθεια είναι 9.925 €.

α) Ο αριθμός $\hat{y}_0 = 9.925$ είναι μια συστηματική εκτίμηση των κερδών αν στην πραγματικότητα το x είναι το 2,5. Αυτά, $\epsilon(y|x=2,5)$

Για να δούμε το αντίστοιχο Α.Ε. υπολογίζουμε:

$$\sum x_i^2 - n \cdot \bar{x}^2 = 20$$

$$(x_0 - \bar{x})^2 = (2,5 - 3)^2 = (-0,5)^2 = 0,25$$

$$s_e = \sqrt{\frac{\sum (y_i^2 - n \cdot \bar{y}^2) - \hat{\beta}_1^2 (\sum x_i^2 - n \cdot \bar{x}^2)}{n-2}} = \sqrt{\frac{114 - 2,15^2 \cdot 20}{8}} \Rightarrow$$

$$s_e = \sqrt{\frac{114 - 22,45}{8}} = \sqrt{\frac{91,55}{8}} = \sqrt{2,6375} \approx 1,64$$

$$t_{0,12}(8) = 1,86$$

Το Δ.Ε. είναι:

$$(2,925 \pm 1,86 \cdot 1,64 \sqrt{\frac{1}{10} + \frac{0,25}{20}})$$

$$(2,925 \pm 1,86 \cdot 1,64 \sqrt{0,1 + 0,0125})$$

$$(2,925 \pm 1,86 \cdot 1,64 \cdot 0,33)$$

$$(2,925 \pm 1,006)$$

$$(3,919, 10,931)$$

Διλαδή, η άγνωστη $E(Y|X=2,5)$ βρίσκεται στο διάστημα $(3,919, 10,931)$ με πιθανότητα 90%.

β)

Η συνάρτηση εκτίμησης του Y_0 όταν $X_0 = 2,5$ ενώ το $\hat{Y}_0 = 2,925$ κλ. €

Το Δ.Ε. είναι:

$$(2,925 \pm 1,86 \cdot 1,64 \sqrt{1 + \frac{1}{10} + \frac{0,25}{20}})$$

$$(2,925 \pm 1,86 \cdot 1,64 \sqrt{1,1125})$$

$$(2,925 \pm 3,20)$$

$$(6,725, 13,125)$$

Διλαδή, τα κέρδη για κάποια εταιρία που βάζει 2.500 για διαφήμιση είναι από 6.725 € έως 13.125 € με πιθανότητα 90%.

Άσκηση 21

Ενδιαφερόμαστε για την συσχέτιση ετήσιων απολαβών των υπαλλήλων μιας εταιρίας με σχέση με την προώθηση τους στην εταιρία. Στον παρακάτω πίνακα δίνεται:

x_i : αριθμός ετών προώθησης

y_i : ετήσιος μισθός σε €.

Δεκα υπαλλήλων της εταιρίας.

											Σύνολο
x_i	1	2	2	3	4	6	9	10	10	13	60
y_i	12	13	14	17	16	17	18	20	20	23	170

Να υπολογίσετε:

i) Pearson

ii) Να βρεθεί η ευθεία ελαχίστων τετραγώνων του μισθού στα έτη προώθησης

iii) Να δώσετε συμπεριλαβή εκτίμηση

a) Για το μέσο ετήσιο μισθό υπαλλήλων που έχουν 8 χρόνια προώθησης.

b) Για τον ετήσιο μισθό ενός υπαλλήλου με 8 χρόνια προώθησης.

iv) Να δώσετε ένα 95% Διαστήμα Εμπιστοσύνης:

a) Για το μέσο ετήσιο μισθό υπαλλήλων που έχει 8 έτη προώθησης.

b) Για τον ετήσιο μισθό ενός υπαλλήλου από τον πληθυσμό που έχει 12 έτη προώθησης.

Λύση:

$$\bar{x} = \frac{60}{10} = 6, \quad \bar{y} = \frac{170}{10} = 17$$

											Σύνολο
$x_i y_i$	12	26	28	51	64	102	162	200	200	239	1.444
x_i^2	1	4	4	9	16	36	81	100	100	169	520
y_i^2	144	169	196	289	256	289	324	400	400	529	2.996

i)

$$r = \frac{\sum x_i y_i - v \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum x_i^2 - v \cdot \bar{x}^2} \sqrt{\sum y_i^2 - v \cdot \bar{y}^2}} \Rightarrow r = \frac{1144 - 10 \cdot 6 \cdot 17}{\sqrt{520 - 10 \cdot 6^2} \sqrt{2996 - 10 \cdot 17^2}} \Rightarrow$$

$$r = \frac{1144 - 1020}{\sqrt{520 - 360} \sqrt{2996 - 2890}} \Rightarrow r = \frac{124}{\sqrt{160} \sqrt{106}} \Rightarrow r = \frac{124}{\sqrt{16960}} \Rightarrow$$

$$r = \frac{124}{130,23} \Rightarrow r = 0,95216$$

Υπάρχει πολύ ισχυρή γραμμική συσχέτιση.

ii)

$$\hat{b}_1 = \frac{\sum x_i y_i - v \cdot \bar{x} \cdot \bar{y}}{\sum x_i^2 - v \cdot \bar{x}^2} \Rightarrow \hat{b}_1 = \frac{124}{60} \Rightarrow \hat{b}_1 = 0,775$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \cdot \bar{x} \Rightarrow \hat{b}_0 = 17 - 0,775 \cdot 6 \Rightarrow \hat{b}_0 = 17 - 4,65 \Rightarrow \hat{b}_0 = 12,35$$

$$\hat{y} = \hat{b}_0 + \hat{b}_1 \cdot x$$

$$\hat{y} = 12,35 + 0,775 \cdot x$$

iii) Για $x=8$ μετρία που δίνει:

$$a) \hat{y} = 12,35 + 0,775 \cdot 8 \Rightarrow$$

$$\hat{y} = 12,35 + 6,2 \Rightarrow$$

$$\hat{y} = 18,55 \text{ πιν. € ή } 18.550 \text{ f.}$$

β) Ο εκτιμώμενος μέσος ενός υπαλλήλου για 8 χρόνια προϋπηρεσία είναι 18.550 € αφού το έχουμε υπολογίσει συστηματικά για όσους έχουν 8 χρόνια, άρα και ο ένας υπάλληλος θα έχει τόσα χρήματα.

iv)

α) Ζητάμε Δ.Ε. 95 % για την $E(Y/X=8)$

$$\bullet \hat{Y}_0 = 18,55 \text{ χιλ. €}$$

$$\bullet \sum (y_i - \bar{y})^2 = \sum y_i^2 - n \cdot \bar{y}^2 = 106$$

$$\bullet \sum (x_i - \bar{x})^2 = \sum x_i^2 - n \cdot \bar{x}^2 = 160$$

$$\bullet Se = \sqrt{\frac{106 - 0,775^2 \cdot 160}{8}} \Rightarrow Se = \sqrt{\frac{106 - 0,600 \cdot 160}{8}} \Rightarrow Se = \sqrt{\frac{106 - 0,61}{8}}$$

$$Se = \sqrt{\frac{9,9}{8}} \Rightarrow Se = \sqrt{1,2375} \Rightarrow Se \approx 1,11$$

$$\bullet \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} = \frac{1}{10} + \frac{(8 - 6)^2}{160} = \frac{1}{10} + \frac{2^2}{160} = \frac{1}{10} + \frac{4}{160} = 0,125$$

Δ.Ε.

$$(18,55 \pm 2,306 \cdot 1,11 \sqrt{0,125})$$

$$(18,55 \pm 2,306 \cdot 1,11 \cdot 0,3535)$$

$$(18,55 \pm 0,90)$$

$$(17,65, 19,45)$$

Με πιθανότητα 95 % $q \in (Y/X=8)$ είναι από 17,65 έως 19,45 χιλ. €

Άσκηση (22)

Ένα τυχαίο δείγμα 10 δανειοληπτών επιλέγεται τυχαία και εξετάζεται ως προς τις μεταβλητές:

X : πλήθος δανείων που έχουν λάβει

Y : αριθμός διαφορετικών τραπεζών από τις οποίες τα έχουν πάρει.

Οι απαντήσεις του δείγματος φαίνονται παρακάτω:

x_i	1	2	2	3	3	3	3	4	4	5	Σύνολο
y_i	1	1	2	1	1	2	3	2	3	4	30
											20

Δίνονται: $\sum x_i^2 = 102$, $\sum y_i^2 = 50$

α) Να εξετασθεί αν υπάρχει γραμμική συσχέτιση

β) Να εκτιμήσετε την ευθεία παλινδρόμησης της Y στην X (να βρείτε την ευθεία ελαχίστων τετραγώνων).

Λύση:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{30}{10} = 3, \quad \bar{y} = \frac{\sum y_i}{n} = \frac{20}{10} = 2$$

$x_i y_i$	1	2	4	3	3	6	9	8	12	20	Σύνολο
											68

$$r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum x_i^2 - n \cdot \bar{x}^2} \sqrt{\sum y_i^2 - n \cdot \bar{y}^2}} \Rightarrow r = \frac{68 - 10 \cdot 3 \cdot 2}{\sqrt{102 - 10 \cdot 3^2} \sqrt{50 - 10 \cdot 2^2}} \Rightarrow r = \frac{68 - 60}{\sqrt{102 - 90} \sqrt{50 - 40}}$$

$$r = \frac{8}{\sqrt{12} \sqrt{10}} \Rightarrow r = \frac{8}{3,46 \cdot 3,16} \Rightarrow r = \frac{8}{10,93} \Rightarrow r \approx 0,73$$

β)

$$\hat{b}_1 = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum x_i^2 - n \cdot \bar{x}^2} \Rightarrow \hat{b}_1 = \frac{8}{12} \Rightarrow \hat{b}_1 = 0,66$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \cdot \bar{x} \Rightarrow \hat{b}_0 = 2 - 0,66 \cdot 3$$

$$\Rightarrow \hat{b}_0 = 2 - 1,98$$

$$\Rightarrow \hat{b}_0 = 0,02$$

Άρα, η ευθεία ελαχίστων τετραγώνων της Y στην X είναι:

$$\hat{Y} = 0,02 + 0,66 X$$

Άσκηση (23)

Ενδιαφερόμαστε για την συσχέτιση του αριθμού των διαφημίσεων των ηλεκτρονικών συσκευών μιας εταιρίας με τις πωλήσεις αυτών των συσκευών πανελλαδικά. Στον παρακάτω πίνακα δίνεται:

x_i : αριθμός διαφημίσεων / μήνα (σε δεκάδες)

y_i : μηνιαίες πωλήσεις αυτών των συσκευών (σε δεκάδες)

	Συνολο									
x_i	3	1	4	0	2	4	0	3	1	2
y_i	26	12	30	6	20	33	5	28	15	25
	20									
	200									

Δίνεται: $\sum x_i y_i = 531$, $\sum y_i^2 = 4.904$

α) Να προσδιοριστεί το είδος της γραμμικής συσχέτισης των δυο μεταβλητών

(3)

ii) Να βρεθεί η ευθεία ελαχίστων τετραγώνων της Y πάνω στην X .

iii) Να δώσετε γραφική εκτίμηση:

α) Για τον κατά μέσο μηνιαίο αριθμό πωλήσεων συσκευών αν πραγματοποιηθούν 45 διαφημίσεις κατά την διάρκεια του μήνα

β) Για τον αριθμό των πωλήσεων του επόμενου μήνα αν η εταιρία αποφασίσει να πραγματοποιήσει 45 διαφημίσεις.

iv) Να δώσετε ένα 95% Διαστήμα Εμπιστοσύνης.

α) Για τον μέσο μηνιαίο αριθμό πωλήσεων συσκευών αν πραγματοποιηθούν 45 διαφημίσεις κατά τη διάρκεια του μήνα.

Λύση:

$$\bar{x} = \frac{20}{10} = 2$$

$$\bar{y} = \frac{200}{10} = 20$$

x_i^2	9	1	16	0	4	16	0	9	1	4	Σύνολο
											60

i)

$$r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum x_i^2 - n \cdot \bar{x}^2} \sqrt{\sum y_i^2 - n \cdot \bar{y}^2}} \Rightarrow r = \frac{531 - 10 \cdot 2 \cdot 20}{\sqrt{60 - 10 \cdot 2^2} \sqrt{4204 - 10 \cdot 20^2}} \Rightarrow$$

$$r = \frac{531 - 400}{\sqrt{60 - 40} \sqrt{4204 - 4000}} \Rightarrow r = \frac{131}{\sqrt{20} \sqrt{904}} \Rightarrow r = \frac{131}{4,47 \cdot 30,06} \Rightarrow$$

$$r = \frac{131}{134,36} \Rightarrow r = 0,97$$

ii)

$$\hat{b}_1 = \frac{\sum x_i y_i - v \cdot \bar{x} \cdot \bar{y}}{\sum x_i^2 - v \cdot \bar{x}^2} \Rightarrow \hat{b}_1 = \frac{131}{20} \Rightarrow \hat{b}_1 = 6,55$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \cdot \bar{x} \Rightarrow \hat{b}_0 = 20 - 6,55 \cdot 2 \Rightarrow \hat{b}_0 = 20 - 13,1 \Rightarrow \hat{b}_0 = 6,9$$

$$\boxed{\hat{Y} = 6,9 - 6,55x}$$

iii) Για $x = 4,5$

$$\hat{Y} = 6,9 + 6,55 \cdot 4,5$$

$$\hat{Y} = 6,9 + 29,475$$

$$\hat{Y} = 36,375 \quad \text{ήδη ευρήματα για τα επείγοντα α) και β)}$$

iv) Ζητάμε Α.Ε. ^{95%} για τον $E(Y|X=4,5)$

$$\bullet Se = \sqrt{\frac{\sum (y_i - \bar{y})^2 - \hat{b}_1^2 \sum (x_i - \bar{x})^2}{v-2}} \Rightarrow Se = \sqrt{\frac{304 - 6,55^2 \cdot 20}{8}} \Rightarrow$$

$$Se = \sqrt{5,73} \Rightarrow Se \approx 2,39$$

$$\bullet \sqrt{\frac{1}{v} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = \sqrt{0,1 + \frac{(4,5 - 2)^2}{20}} = \sqrt{0,1 + \frac{6,25}{20}} = \sqrt{0,1 + 0,3125}$$

$$= \sqrt{0,4125} \approx 0,64$$

$$\bullet t_{0,025}(2) = 2,306$$

Αντικαθιστώντας στο τύπο και έχουμε :

$$(36,375 \pm 2,306 \cdot 2,32 \cdot 0,64)$$

$$(36,375 \pm 3,52)$$

$$(32,855, 39,895)$$

Με πιθανότητα 95 % ο αριθμός $\epsilon(Y/X=4,5)$ είναι 32,855 έως 39,895.
 Δηλαδή με 45 διαφημίσεις περιμένω να πουλήσω (32,855, 39,895) με 95%.

Άσκηση (24)

Το υπουργείο συγκοινωνιών μελέτησε τον αριθμό των θανατηφόρων ατυχημάτων με δίτροχα που οφείλονται στην μέθη τα τελευταία 8 χρόνια.

Τα στοιχεία φαίνονται παρακάτω:

X_i	Y_i
1985	190
1986	185
1987	195
1988	180
1989	200
1990	185
1991	190
1992	205

i) Να βρείτε την ευθεία ελαχίστων τετραγώνων και να προβλέψετε τον αριθμό των ατυχημάτων το 1993.