

Παράδειγμα 4 (Φροντιστήριο 1)

Δίνονται τα ακόλουθα δεδομένα μίας επιχείρησης από τις πωλήσεις (Y) και τη διαφημιστική δαπάνη (X) για τα έτη 1985-2000. Επίσης δίνονται και οι απαραίτητοι υπολογισμοί για να προσδιορισμό της τιμής του δειγματικού συντελεστή συσχέτισης r των X και Y.

Ετος	Πωλήσεις (Y)	Διαφήμιση (X)	YX	Y ²	X ²
1985	1050	162	170100	1102500	26244
1986	1260	285	359100	1587600	81225
1987	1470	540	793800	2160900	291600
1988	2160	261	563760	4665600	68121
1989	1950	360	702000	3802500	129600
1990	2400	690	1656000	5760000	476100
1991	2370	495	1173150	5616900	245025
1992	3150	948	2986200	9922500	898704
1993	3570	720	2570400	12744900	518400
1994	4410	1140	5027400	19448100	1299600
1995	4500	1395	6277500	20250000	1946025
1996	5610	1560	8751600	31472100	2433600
1997	5190	1380	7162200	26936100	1904400
1998	5670	1260	7144200	32148900	1587600
1999	5160	1710	8823600	26625600	2924100
2000	6840	1860	12722400	46785600	3459600
Άθροισμα	56760	14766	66883410	251029800	18289944

$$r = \frac{16 (66.883.410) - (56.760) (14.766)}{\sqrt{(16 (18.289.944) - (14.766)^2) (16 (251.029.800) - (56.760)^2)}} = 0.95$$

Επίσης δίνονται τα ακόλουθα δεδομένα μίας επιχείρησης από τις πωλήσεις (Y) και τους πωλητές (X) για τα ίδια έτη 1985-2000. Επίσης δίνονται και οι απαραίτητοι υπολογισμοί για τον προσδιορισμό της τιμής του δειγματικού συντελεστή συσχέτισης r των X και Y.

Ετος	Πωλήσεις (Y)	Πωλητές (X)	YX	Y ²	X ²
1985	1050	32	33600	1102500	1024
1986	1260	47	59220	1587600	2209
1987	1470	23	33810	2160900	529
1988	2160	68	146880	4665600	4624
1989	1950	32	62400	3802500	1024
1990	2400	17	40800	5760000	289
1991	2370	58	137460	5616900	3364
1992	3150	75	236250	9922500	5625
1993	3570	98	349860	12744900	9604
1994	4410	43	189630	19448100	1849
1995	4500	76	342000	20250000	5776
1996	5610	89	499290	31472100	7921
1997	5190	108	560520	26936100	11664
1998	5670	76	430920	32148900	5776
1999	5160	65	335400	26625600	4225
2000	6840	93	636120	46785600	8649
Άθροισμα	56760	1000	4094160	251029800	74152

$$r = \frac{16(4.094.160) - (56.760)(1.000)}{\sqrt{(16(74.152) - (1.000)^2)(16(251.029.800) - (56.760)^2)}} = 0.72$$

Πριν το υπολογισμό του r σε κάθε περίπτωση, ο πιο απλός τρόπος για να διαπιστώσουμε αν υπάρχει συσχέτιση μεταξύ δύο μεταβλητών είναι η κατασκευή του διαγράμματος διασποράς. Όπως βλέπουμε από τα σχήματα, το διάγραμμα διασποράς μεταξύ των πωλήσεων και διαφήμισης αποκαλύπτει τη θετική σχέση μεταξύ των 2 μεταβλητών. Το διάγραμμα διασποράς μεταξύ των πωλητών και διαφήμισης αποκαλύπτει τη θετική σχέση μεταξύ των 2 μεταβλητών, η οποία δεν είναι τόσο έντονη όσο στη προηγούμενη περίπτωση.

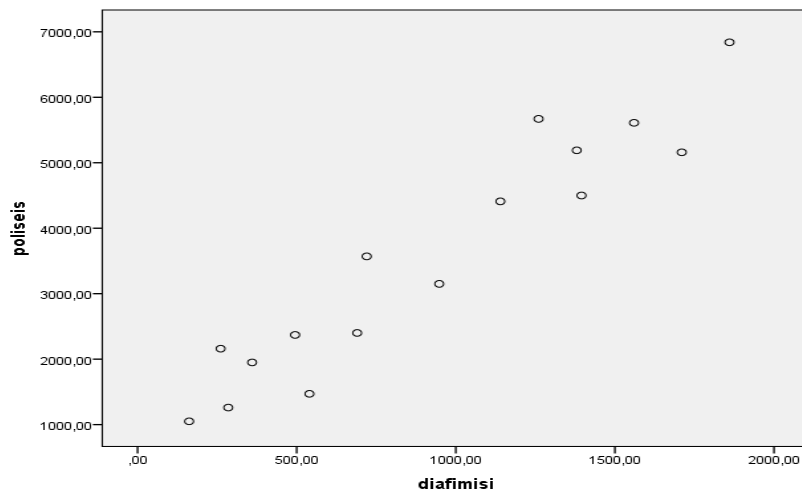
Ο συντελεστής συσχέτισης μεταξύ διαφήμισης και πωλήσεων είναι 0.95 και μεταξύ πωλητών και διαφήμισης είναι 0.72. Ο έλεγχος στατιστικής σημαντικότητας του πληθυσμιακού συντελεστή συσχέτισης ρ με το κριτήριο t και στις δύο περιπτώσεις μας οδηγεί σε απόρριψη της μηδενικής υπόθεσης για ε.σ.σ. $\alpha = 0.05$.

Ας κάνουμε κάποια τελευταία σχόλια για το συντελεστή συσχέτισης. Τα αποτελέσματα των ελέγχων παρέχουν ενδείξεις γραμμικής σχέσης-συσχέτισης μεταξύ των δύο μεταβλητών X και Y σε κάθε περίπτωση. Μια χαμηλή τιμή του συντελεστή συσχέτισης δεν σημαίνει ότι δεν υπάρχει γραμμική συσχέτιση, γιατί μπορεί να είναι ασθενής και μπορεί να υπάρχει άλλου είδους συσχέτιση (π.χ. καμπυλόγραμμη).

Μπορεί οι μεταβλητές, για παράδειγμα $X :=$ πωλήσεις αυτοκινήτων και $Y :=$ πωλήσεις πίτσας να είναι υψηλά συσχετισμένες, αλλά αυτό να οφείλεται ότι και οι δύο είναι συσχετισμένες με ένα τρίτο παράγοντα (αλλαγή τρόπου διαβίωσης). Σε μια τέτοια περίπτωση η συσχέτιση αυτή καλείται **νόθος συσχέτιση (spurious correlation)**.

Διαγράμματα Διασποράς

A. Διαφήμιση (X)-Πωλήσεις (Y)



B. Πωλητές (X)-Πωλήσεις (Y)

