

Υπολογισμός του Δειγματικού Συντελεστή Συσχέτισης

Ο δειγματικός συντελεστής συσχέτισης είναι ο μέσος των γινομένων των τυποποιημένων δειγματικών τιμών των ζευγών x και y των μεταβλητών X και Y .

Ο μαθηματικός τύπος που δίνει τον δειγματικό συντελεστή συσχέτισης είναι

$$r = \frac{\text{Cov}(x, y)}{s_x s_y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$

Όπως έχει αναφερθεί, ο λόγος που χρησιμοποιούμε τις τυποποιημένες τιμές των X και Y είναι για να αποφύγουμε τις δυσκολίες που προκαλούνται, όταν τα μεγέθη την σχέση των οποίων θέλουμε να προσδιορίσουμε μετρώνται σε διαφορετικές κλίμακες. (Όπως είναι γνωστό, ο τρόπος για να αποφεύγεται το πρόβλημα αυτό είναι η τυποποίηση των τιμών. Η τυποποιημένη τιμή μιας τυχαίας μεταβλητής (που προκύπτει αν από την τιμή αφαιρέσουμε την μέση τιμή των παρατηρήσεων και την διαφορά την διαιρέσουμε με την τυπική απόκλιση των παρατηρήσεων) δείχνει, για κάθε τιμή, πόσες φορές η τιμή αυτή είναι πάνω ή κάτω από την μέση τιμή σε σχέση με την τυπική απόκλιση).

Η διαδικασία, δηλαδή, για τον υπολογισμό του δειγματικού συντελεστή συσχέτισης είναι ο μετασχηματισμός των τιμών κάθε μεταβλητής στις αντίστοιχες τυποποιημένες τιμές τους και στην συνέχεια ο υπολογισμός του μέσου των γινομένων των τυποποιημένων αυτών τιμών.

Αριθμητικό παράδειγμα: Να υπολογισθεί ο δειγματικός συντελεστής συσχέτισης r για τα δεδομένα που ακολουθούν.

<u>x</u>	<u>y</u>
1	5
3	9
4	7
5	1
7	13

Παρατήρηση: Η πρώτη γραμμή του παραπάνω πίνακα αποτυπώνει δύο μετρήσεις του ίδιου αντικειμένου στην συγκεκριμένη μελέτη. Οι δύο αριθμοί είναι οι x και y συντεταγμένες του αντιστοίχου σημείου στο διάγραμμα (νέφος) σημείων. Το ίδιο συμβαίνει και με τις υπόλοιπες γραμμές. Έχει σημασία να χρησιμοποιηθούν οι αριθμοί του πίνακα ως ζεύγη. Ο συντελεστής συσχέτισης ορίζεται μόνο όταν έχουμε δύο μεταβλητές οι οποίες μετρώνται για κάθε αντικείμενο της μελέτης.

Βήμα 1: Μετατρέπουμε τις τιμές της μεταβλητής X σε τυποποιημένες τιμές. Για να γίνει αυτό χρειαζόμαστε τον μέσο \bar{X} και την τυπική απόκλιση s_x των τιμών του x .

Βρίσκουμε ότι $\bar{X} = 4$, $s_x = 2$.

Στην συνέχεια, αφαιρούμε τον μέσο από κάθε τιμή της μεταβλητής X και διαιρούμε με την τυπική απόκλιση.

$$\frac{1-4}{2} = -1.5 \quad \frac{3-4}{2} = -0.5 \quad \frac{4-4}{2} = 0 \quad \frac{5-4}{2} = 0.5 \quad \frac{7-4}{2} = 1.5$$

Οι αριθμοί αυτοί προσδιορίζουν πόσο μακριά, πάνω ή κάτω, βρίσκονται οι τιμές της μεταβλητής X από τον μέσο όρο τους σε σχέση με την τυπική απόκλιση. Στο παράδειγμά, μας η τιμή 1 είναι 1.5 τυπικές αποκλίσεις χαμηλότερη από τον μέσο.

Βήμα 2: Μετατρέπουμε τις τιμές της μεταβλητής Y σε τυπικές τιμές.

Βήμα 3: Για κάθε ένα από τα ζευγάρια των τυπικών τιμών υπολογίζουμε το γινόμενο: (τυπική τιμή του x) * (τυπική τιμή του y).

Υπολογισμός του Συντελεστή Συσχέτισης r

x	y	Τυποποιημένες τιμές του x	Τυποποιημένες τιμές του y	Γινόμενο
1	5	-1.5	-0.5	0.75
3	9	-0.5	0.5	-0.25
4	7	0.0	0.0	0.00
5	1	0.5	-1.5	-0.75
7	13	1.5	1.5	2.25

Βήμα 4: Θεωρούμε τον μέσο των γινομένων

$$\begin{aligned} r &= \text{μέσος (τυπικών τιμών του x)} * \text{(τυπικών τιμών του y)} = \\ &= \frac{0.75 - 0.25 + 0.00 - 0.75 + 2.25}{5} = 0.40 \end{aligned}$$

