Ο πίνακας συσχέτισης (Correlation matrix) R

June 4, 2021

Ο πίνακας συσχέτισης

Ο πίνακας συσχέτισης R είναι ενας τετραγωνικός πίνακας που δείχνει τις συσχετίσεις (τον συντελεστή συσχέτισης) μεταξύ των διάφορων μεταβλητών. Τα διαγώνια στοιχεία του είναι ισα με 1. Είναι χρήσιμος στην ανάλυση πολυμεταβλητών δεδομένων, και δείχνει τους συντελεστές συσχέτισης μεταξύ τους. Είναι άμεσα συνδεδεμένος με τον πίνακα διακύμανσης - συνδιακύμανσης Σ που έχουμε ηδη αναλύσει σε προηγούμενο μάθημα.
Ο πίνακας συσχέτισης των ν τυχαίων μεταβλητών $X_1, \ldots, X_n$ είναι ο $n \times n$ πίνακας του οποίου τα i, j της εισόδου είναι $corr(X_i, X_j)$. Αν τα μέτρα της συσχέτισης χρησιμοποιούνται ως συντελεστές, ο πίνακας συσχέτισης είναι ο ίδιος με τον πίνακα συνδιασποράς των τυποποιημένων τυχαίων μεταβλητών $X_i/\sigma(X_i)$ για $i = 1, \ldots n$. Αυτό ισχύει τόσο για τον πίνακα συσχετίσεων του πληθυσμού (στην περίπτωση αυτή το σ είναι η τυπική απόκλιση πληθυσμού), οσο για τον πίνακα συσχετίσεων του δείγματος (στην περίπτωση αυτή το σ, δηλώνει η τυπική απόκλιση του δείγματος). Κατά συνέπεια, το καθένα είναι απαραιτήτως ένας θετικός-ημιθετικός πίνακας.
Ο πίνακας συσχετίσεων είναι συμμετρικος διότι η συσχέτιση μεταξύ των X_i, X_j είναι η ίδια όπως η συσχέτιση ανάμεσα σε X_j, X_i.
Η σχέση μεταξύ του πίνακα συσχέτισης R και του πίνακα διακύμανσης- συνδιακύμανσης είναι η ακόλουθη:

$$R = \Delta^{-\frac{1}{2}} \Sigma \Delta^{-\frac{1}{2}}$$

ισοδύναμα:

$$\Sigma = \Delta^{\frac{1}{2}} R \Delta^{\frac{1}{2}}$$

όπου Δ ένας διαγώνιος πίνακας που έχει στην διαγώνιο του τα αντίστοιχα στοιχεία του πίνακα Σ. (τις $Var(X_i)$) Τα διαγώνια στοιχεία του πίνακα συσχέτισης είναι μονάδες.

Παράδειγμα 1 *Έστω ο πίνακας διακύμανσης δύο μεταβλητών Σ.*

$$\Sigma = \begin{bmatrix} 1 & 1 & 8.1 \\ 1 & 16 & 18 \\ 8.1 & 18 & 81 \end{bmatrix}$$

Να υπολογιστεί ο πίνακας συσχέτισης R.
Λύση:
Βλέπουμε οτι η διαγώνιος του πίνακα Σ αποτελείται απο τους αριθμούς $1, 16, 81$, *αρα ο πίνακας Δ θα είναι ο*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 81 \end{bmatrix} \quad \text{αρα} \quad \mathbf{\Delta}^{-\frac{1}{2}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{9} \end{bmatrix}$$

Ετσι παίρνουμε:

$$R = \Delta^{-\frac{1}{2}} \Sigma \Delta^{-\frac{1}{2}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{9} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 81 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{9} \end{bmatrix}$$

Ετσι έχουμε οτι

$$R = \begin{bmatrix} 1 & 0.25 & 0.9 \\ 0.25 & 1 & 0.5 \\ 0.9 & 0.5 & 1 \end{bmatrix}$$

Στην περίπτωση του 2×2 πίνακα R, της μορφής

$$R = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$$

οι ιδιοτιμές του είναι $1 - r, 1 + r$. Τα αντίστοιχα ιδιοδιανύσματα είναι $(1, 1)$ και $(-1, 1)$.

$\triangleleft$