

# Πολυμεταβλητή Ανάλυση

## Τυχαία διανύσματα

**Λουκία Μελιγκοτσίδου**

**Τμήμα Μαθηματικών, ΕΚΠΑ**

## 1 Τυχαία διανύσματα

- Τυχαίο διάνυσμα είναι μια συλλογή (διάνυσμα) τυχαίων μεταβλητών.
- Συμβολίζεται ως στήλη ή γραμμή

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \text{ ή } \mathbf{X}^\top = (X_1, X_2, \dots, X_p)$$

- Εξετάζουμε την από κοινού κατανομή τους (π.χ. από κοινού κατανομή ύψους, βάρους, ηλικίας)
- Κάθε στοιχείο του  $\mathbf{X}$  είναι μια τυχαία μεταβλητή η οποία έχει μια **περιθώρια** κατανομή.

## Θεμελιώδης έννοια

- η από κοινού συνάρτηση πυκνότητας πιθανότητας (σππ) (joint probability density function)  $f(x_1, x_2, \dots, x_p)$  για τις συνεχείς τυχαίες μεταβλητές
- η από κοινού συνάρτησης πιθανότητας (σπ) (joint probability function)  $P(X_1 = x_1, X_2 = x_2, \dots, X_p = x_p)$  για την περίπτωση διακριτών τυχαίων μεταβλητών

# Περιθώριες κατανομές (συνεχείς τυχαίες μεταβλητές)

Περιθώρια κατανομή μιας τμ (έστω  $X_1$ )

$$f(x_1) = \int_{x_2} \int_{x_3} \dots \int_{x_p} f(x_1, x_2, \dots, x_p) dx_p \dots dx_2$$

Από κοινού περιθώρια κατανομή δύο ή περισσότερων τμ (έστω  $X_1, X_2$ )

$$f(x_1, x_2) = \int_{x_3} \dots \int_{x_p} f(x_1, x_2, \dots, x_p) dx_p \dots dx_3$$

Στη γενική περίπτωση, από κοινού περιθώρια των τμ ( $m < p$ )  $X_1, X_2, \dots, X_m$

$$f(x_1, \dots, x_m) = \int_{x_{m+1}} \int_{x_{m+2}} \dots \int_{x_p} f(x_1, x_2, \dots, x_p) dx_p \dots dx_{m+1}$$

## Κανόνας

Γενικά για να βρούμε μια από κοινού περιθώρια κατανομή ολοκληρώνουμε ως προς όλες τις μεταβλητές που θέλουμε να μην εμφανίζονται μέσα στην από κοινού σππ

# Δεσμευμένες κατανομές

Δεσμευμένη κατανομή διανύσματος (έστω  $X_1, X_2$ ) δοθέντων των τιμών άλλων τμ (έστω  $X_3, \dots, X_p$ )

$$f(x_1, x_2 | x_3, \dots, x_p) = \frac{f(x_1, x_2, \dots, x_p)}{f(x_3, \dots, x_p)}$$

Αποτελεί τη γενίκευση του γνωστού τύπου από πιθανότητες

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Απαντά στο ερώτημα, π.χ για δεδομένο ύψος 180cm, ποια η κατανομή του βάρους;
- Εάν κρατήσουμε όλα τα άτομα με ύψος 180cm, ποιά θα είναι η κατανομή πιθανότητας του βάρους;

## Αναμενόμενη τιμή για οποιαδήποτε συνάρτηση τιμ

$$E\{g(X_1, X_2, \dots, X_p)\} = \int_{x_1} \dots \int_{x_p} g(x_1, x_2, \dots, x_p) f(x_1, x_2, \dots, x_p) dx_p \dots dx_1$$

Ειδική περίπτωση,  $g(X_1, X_2, \dots, X_p) = X_1$

$$\begin{aligned} E(X_1) &= \int_{x_1} \dots \int_{x_p} x_1 f(x_1, x_2, \dots, x_p) dx_p \dots dx_1 \\ &= \int_{x_1} x_1 \left\{ \int_{x_2} \dots \int_{x_p} f(x_1, x_2, \dots, x_p) dx_p \dots dx_2 \right\} dx_1 \\ &= \int_{x_1} x_1 f(x_1) dx_1 \end{aligned}$$

η αναμενόμενη τιμή της περιθώριας κατανομής της  $X_1$ .

Οι τμ  $X_1, X_2, \dots, X_p$  καλούνται από κοινού ανεξάρτητες εάν η από κοινού κατανομή τους γράφεται ως

$$f(x_1, x_2, \dots, x_p) = f(x_1)f(x_2) \dots f(x_p)$$

για κάθε  $(x_1, x_2, \dots, x_p)^T \in \mathbb{R}^p$ .

# Αναμενόμενες τιμές τυχαίου διανύσματος

- Έστω  $\mathbf{X} = (X_1, X_2, \dots, X_p)$  τυχαίο διάνυσμα
- Αναμενόμενη τιμή  $\mathbf{X}$

$$\boldsymbol{\mu} = E(\mathbf{X}) = \begin{Bmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{Bmatrix}_{p \times 1}$$

- Έστω  $\mathbf{c} \in \mathbb{R}^p$  σταθερό διάνυσμα, τότε η τμ  $Y = \mathbf{c}^\top \mathbf{X}$  έχει μέση τιμή  $E(Y) = \mathbf{c}^\top \boldsymbol{\mu}$

$$Y = c_1 X_1 + \dots + c_p X_p \Rightarrow E(Y) = c_1 \mu_1 + \dots + c_p \mu_p = \mathbf{c}^\top \boldsymbol{\mu}$$

- Αναλόγως, εάν  $\mathbf{Y} = \mathbf{C}\mathbf{X}$ , όπου  $\mathbf{C}_{m \times p}$  σταθερός πίνακας, τότε  $\mathbf{Y}_{m \times 1}$  τυχαίο διάνυσμα με  $E(\mathbf{Y}) = \mathbf{C}\boldsymbol{\mu}$ .

Ο πίνακας  $\mathbf{X}$  διαστάσεων  $m \times p$  του οποίου όλα τα στοιχεία είναι τυμ λέγεται τυχαίος πίνακας

- Αναμενόμενη τιμή τυ

$$E(\mathbf{X}) = \begin{Bmatrix} E(X_{11}) & E(X_{12}) & \vdots & E(X_{1p}) \\ E(X_{21}) & E(X_{22}) & \vdots & E(X_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_{m1}) & E(X_{m2}) & \vdots & E(X_{mp}) \end{Bmatrix}_{m \times p}$$

Έστω  $\mathbf{X}$  και  $\mathbf{Y}$  τυχαίοι πίνακες ίδιων διαστάσεων  $\Rightarrow$

$$E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$$

Επίσης, εάν  $\mathbf{A}$  και  $\mathbf{B}$  πίνακες κατάλληλων διαστάσεων (όχι όμως τυχαίοι)

$$E(\mathbf{AXB}) = \mathbf{AE(X)B}$$

# Πίνακας συνδιακύμανσης τυχαίου διανύσματος $\mathbf{X}$

Ορίζεται ως

$$\text{Var}(\mathbf{X}) = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_2, X_1) & \vdots & \text{Cov}(X_p, X_1) \\ & \text{Var}(X_2) & \vdots & \text{Cov}(X_p, X_2) \\ & & \ddots & \vdots \\ & & & \text{Var}(X_p) \end{pmatrix}_{p \times p}$$

συμμετρικός και θετικά ημιορισμένος

- Η συνάρτηση

$\sigma_{jk} = \text{Cov}(X_j, X_k) = E\{(X_j - \mu_j)(X_k - \mu_k)\} = E(X_j X_k) - E(X_j)E(X_k)$  είναι η συνδιακύμανση μεταξύ  $X_j$  και  $X_k$

- Μέτρο της **γραμμικής** σχέσης μεταξύ δύο μεταβλητών.
- Η συνδιακύμανση μίας τμ με τον εαυτό της είναι η γνωστή μας διακύμανση.
- Όπως είδαμε στην πρώτη διάλεξη, η συνδιακύμανση εκτιμάται από

$$\widehat{\text{Cov}}(X_j, X_k) = s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, X) = Var(X)$
- $Cov(aX + b, cY + d) = acCov(X, Y)$ ,  $a, b, c, d$  σταθερές
- $Cov(X, a) = 0$ ,  $a$  σταθερά
- Εάν  $Cov(X, Y) = 0$  τότε οι  $X, Y$  είναι **ασυσχέτιστες**.

# Πίνακας συνδιακύμανσης τυχαίου διανύσματος $\mathbf{X}$

Ορίζεται ισοδύναμα ως  $\mathbf{\Sigma} = \text{Var}(\mathbf{X}) = E \{ (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top \}$

**Απόδειξη**

$$\begin{aligned} \mathbf{W} &= (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top = \begin{pmatrix} X_1 - \mu_1 \\ \vdots \\ X_p - \mu_p \end{pmatrix} (X_1 - \mu_1 \quad \dots \quad X_p - \mu_p) \\ &= \begin{Bmatrix} (X_1 - \mu_1)^2 & & \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \\ \vdots & \vdots & \ddots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \dots & (X_p - \mu_p)^2 \end{Bmatrix} \end{aligned}$$

Άρα είναι προφανές ότι  $E(\mathbf{W}) = \mathbf{\Sigma}$ . Επίσης είναι εύκολο να δειχθεί ότι

$$\mathbf{\Sigma} = E(\mathbf{X}\mathbf{X}^\top) - \boldsymbol{\mu}\boldsymbol{\mu}^\top.$$

# Ιδιότητες τυχαίων διανυσμάτων

Έστω  $\mathbf{X}_{p \times 1}$  τυχαίο διάνυσμα και  $\mathbf{c}_{p \times 1}$  σταθερό διάνυσμα. Εάν  $Y = \mathbf{c}^\top \mathbf{X}$ ,  $Var(\mathbf{Y}) = \mathbf{c}^\top \mathbf{\Sigma} \mathbf{c}$ , όπου  $\mathbf{\Sigma} = Var(\mathbf{X})$  και  $\boldsymbol{\mu} = E(\mathbf{X})$ .

**Απόδειξη**

$$\begin{aligned} Var(Y) &= Var(\mathbf{c}^\top \mathbf{X}) = E \left\{ (\mathbf{c}^\top \mathbf{X} - \mathbf{c}^\top \boldsymbol{\mu})^2 \right\} = E \left[ \left\{ \mathbf{c}^\top (\mathbf{X} - \boldsymbol{\mu}) \right\}^2 \right] \\ &= E \left\{ \mathbf{c}^\top (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{c} \right\} = \mathbf{c}^\top \mathbf{\Sigma} \mathbf{c} \end{aligned}$$

# Ιδιότητες τυχαίων διανυσμάτων

Έστω  $\mathbf{X}_{p \times 1}$  τυχαίο διάνυσμα και  $\mathbf{c}_{p \times 1}$  σταθερό διάνυσμα. Εάν  $Y = \mathbf{c}^\top \mathbf{X}$ ,  $\text{Var}(Y) = \mathbf{c}^\top \mathbf{\Sigma} \mathbf{c}$ , όπου  $\mathbf{\Sigma} = \text{Var}(\mathbf{X})$  και  $\boldsymbol{\mu} = E(\mathbf{X})$ .

**Απόδειξη**

$$\begin{aligned}\text{Var}(Y) &= \text{Var}(\mathbf{c}^\top \mathbf{X}) = E \left\{ (\mathbf{c}^\top \mathbf{X} - \mathbf{c}^\top \boldsymbol{\mu})^2 \right\} = E \left[ \left\{ \mathbf{c}^\top (\mathbf{X} - \boldsymbol{\mu}) \right\}^2 \right] \\ &= E \left\{ \mathbf{c}^\top (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{c} \right\} = \mathbf{c}^\top \mathbf{\Sigma} \mathbf{c}\end{aligned}$$

## Παρατηρήσεις

- Επειδή  $\text{Var}(Y) = \mathbf{c}^\top \mathbf{\Sigma} \mathbf{c} \geq 0 \rightarrow \mathbf{\Sigma}$  πάντα θετικά ημιορισμένος
- Εάν  $P(\mathbf{c}^\top \mathbf{X} = a) \neq 1 \ \forall \mathbf{c} \in \mathbb{R}^p$  και  $a \in \mathbb{R} \Rightarrow \mathbf{\Sigma}$  θετικά ορισμένος

# Ιδιότητες τυχαίων διανυσμάτων

Έστω  $\mathbf{X}_{p \times 1}$  τυχαίο διάνυσμα και  $\mathbf{c}_{p \times 1}$  σταθερό διάνυσμα. Εάν  $Y = \mathbf{c}^\top \mathbf{X}$ ,  $\text{Var}(Y) = \mathbf{c}^\top \mathbf{\Sigma} \mathbf{c}$ , όπου  $\mathbf{\Sigma} = \text{Var}(\mathbf{X})$  και  $\boldsymbol{\mu} = E(\mathbf{X})$ .

**Απόδειξη**

$$\begin{aligned}\text{Var}(Y) &= \text{Var}(\mathbf{c}^\top \mathbf{X}) = E \left\{ (\mathbf{c}^\top \mathbf{X} - \mathbf{c}^\top \boldsymbol{\mu})^2 \right\} = E \left[ \left\{ \mathbf{c}^\top (\mathbf{X} - \boldsymbol{\mu}) \right\}^2 \right] \\ &= E \left\{ \mathbf{c}^\top (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{c} \right\} = \mathbf{c}^\top \mathbf{\Sigma} \mathbf{c}\end{aligned}$$

## Παρατηρήσεις

- Επειδή  $\text{Var}(Y) = \mathbf{c}^\top \mathbf{\Sigma} \mathbf{c} \geq 0 \rightarrow \mathbf{\Sigma}$  πάντα θετικά ημιορισμένος
- Εάν  $P(\mathbf{c}^\top \mathbf{X} = a) \neq 1 \ \forall \mathbf{c} \in \mathbb{R}^p$  και  $a \in \mathbb{R} \Rightarrow \mathbf{\Sigma}$  θετικά ορισμένος

Γενικότερα, εάν  $\mathbf{Y} = \mathbf{C}\mathbf{X} + \mathbf{b}$ , όπου  $\mathbf{C}_{m \times p}$  και  $\mathbf{b}_{m \times 1}$  γνωστά

$$\text{Var}(\mathbf{Y}) = \mathbf{C}\mathbf{\Sigma}\mathbf{C}^\top$$

Για παράδειγμα, εάν  $\mathbf{Y} = \mathbf{C}\mathbf{X}$ , από τον ορισμό του πίνακα συνδιακύμανσης του τυχαίου διανύσματος  $\mathbf{Y}$

$$\begin{aligned}\text{Var}(\mathbf{Y}) &= E(\mathbf{Y}\mathbf{Y}^\top) - E(\mathbf{Y})E(\mathbf{Y})^\top \\ &= E(\mathbf{C}\mathbf{X}\mathbf{X}^\top\mathbf{C}^\top) - \mathbf{C}\boldsymbol{\mu}\boldsymbol{\mu}^\top\mathbf{C}^\top \\ &= \mathbf{C}\left\{E(\mathbf{X}\mathbf{X}^\top) - \boldsymbol{\mu}\boldsymbol{\mu}^\top\right\}\mathbf{C}^\top = \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top.\end{aligned}$$

# Παράδειγμα 1

Έστω τμ  $X_1$  και  $X_2$ . Ορίζουμε νέες τμ  $Z_1 = X_1 - X_2$  και  $Z_2 = X_1 + X_2$ , εναλλακτικά  $\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$ . Να βρεθούν τα  $E(\mathbf{Z})$  και  $Var(\mathbf{Z})$ .

Γράφουμε  $\mathbf{Z} = \mathbf{C}\mathbf{X}$ , όπου  $\mathbf{C} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$  και  $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ . Άρα

$$E(\mathbf{Z}) = \mathbf{C}\boldsymbol{\mu} = \begin{pmatrix} \mu_1 - \mu_2 \\ \mu_1 + \mu_2 \end{pmatrix}$$

$$\begin{aligned} Var(\mathbf{Z}) &= \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & \sigma_1^2 - \sigma_2^2 \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 + \sigma_2^2 + 2\sigma_{12} \end{pmatrix} \end{aligned}$$

- Τα διαγώνια στοιχεία του πίνακα αποτελούν τα γνωστά μας αποτελέσματα για το άθροισμα και τη διαφορά συσχετισμένων τυχαίων μεταβλητών.
- Προφανώς ο πίνακας που προκύπτει είναι συμμετρικός

## Παράδειγμα 2

Έστω  $\mathbf{Z}^\top = (Z_1, Z_2, Z_3)$ , όπου  $Z_3 = 3X_1 + 4X_2$ , ενώ τα  $Z_1, Z_2$  είναι όπως στο Παράδειγμα 1. Να βρεθούν τα  $E(\mathbf{Z})$  και  $\text{Var}(\mathbf{Z})$ .

Καθώς η  $\mathbf{Z}$  έχει διάσταση  $3 \times 1$ , ο πίνακας  $\mathbf{C}$  πρέπει να έχει διάσταση  $3 \times 2$ .

Άρα,  $\mathbf{C} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 3 & 4 \end{pmatrix}$ , επομένως

$$E(\mathbf{Z}) = \mathbf{C}E(\mathbf{X}) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 3 & 4 \end{pmatrix}_{3 \times 2} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}_{2 \times 1} = \begin{pmatrix} \mu_1 - \mu_2 \\ \mu_1 + \mu_2 \\ 3\mu_1 + 4\mu_2 \end{pmatrix}$$

## Παράδειγμα 2 (συνέχεια)

$$\begin{aligned} \text{Var}(\mathbf{Z}) &= \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 3 & 4 \end{pmatrix}_{3 \times 2} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}_{2 \times 2} \begin{pmatrix} 1 & 1 & 3 \\ -1 & 1 & 4 \end{pmatrix}_{2 \times 3} \\ &= \begin{pmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & \sigma_1^2 - \sigma_2^2 & 3\sigma_1^2 - 4\sigma_2^2 + \sigma_{12} \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 + \sigma_2^2 + 2\sigma_{12} & 3\sigma_1^2 + 4\sigma_2^2 + 7\sigma_{12} \\ 3\sigma_1^2 - 4\sigma_2^2 + \sigma_{12} & 3\sigma_1^2 + 4\sigma_2^2 + 7\sigma_{12} & 9\sigma_1^2 + 16\sigma_2^2 + 24\sigma_{12} \end{pmatrix}_{3 \times 3} \end{aligned}$$

## Παρατήρηση

Οι τύποι που δόθηκαν παραπάνω αφορούν οποιοδήποτε τυχαίο διάνυσμα χωρίς καμία υπόθεση για την από κοινού κατανομή που ακολουθεί το διάνυσμα. Συνεπώς ισχύουν γενικά και σε κάθε περίπτωση

Ορίζουμε τη συσχέτιση δύο τμ  $X_j$  και  $X_k$

$$\rho_{jk} = \rho(X_j, X_k) = \frac{\text{Cov}(X_j, X_k)}{\sigma_j \sigma_k} = \frac{\sigma_{jk}}{\sigma_j \sigma_k}$$

Ισχύουν τα ακόλουθα

- $\rho(X_j, X_k) \in [-1, 1]$
- $\rho(\alpha X_j, \beta X_k) = \rho(X_j, X_k)$  εάν  $\alpha\beta > 0$
- $\rho(\alpha X_j, \beta X_k) = -\rho(X_j, X_k)$  εάν  $\alpha\beta < 0$
- $\rho(X_j, X_k) = 1 \Leftrightarrow X_j = \alpha X_k + b$  με πιθανότητα 1 ( $\alpha > 0$ )
- $\rho(X_j, X_k) = -1 \Leftrightarrow X_j = \alpha X_k + b$  με πιθανότητα 1 ( $\alpha < 0$ )

Εκτιμάται από το δείγμα

$$\hat{\rho}(X_j, X_k) = r_{jk} = \frac{s_{jk}}{s_j s_k} = \frac{\text{Δειγματική συνδιακύμανση}}{\text{Τυπ. απόκλιση } j \times \text{Τυπ. απόκλιση } k}$$

Ο πίνακας συσχετίσεων ενός τυχαίου διανύσματος  $\mathbf{X} = (X_1, X_2, \dots, X_p) \in \mathbb{R}^p$  ορίζεται ως

$$\mathbf{P} = \begin{pmatrix} 1 & \rho_{21} & \dots & \rho_{p1} \\ \rho_{21} & 1 & \dots & \rho_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{pmatrix}$$

- Τα διαγώνια στοιχεία είναι οι συσχετίσεις των μεταβλητών με τον εαυτό τους
- Τα μη διαγώνια στοιχεία είναι οι ανά δύο συσχετίσεις.
- Ο πίνακας  $\mathbf{P}$  είναι συμμετρικός και θετικά ημιορισμένος

# Σχέση πίνακα συσχέτισης και πίνακα συνδιακύμανσης

Ισχύει ότι  $\mathbf{P} = \mathbf{D}^{-1/2} \mathbf{\Sigma} \mathbf{D}^{-1/2}$ , όπου

$$\mathbf{D} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \sigma_p^2 \end{pmatrix}_{p \times p}$$
$$\Rightarrow \mathbf{D}^{-1/2} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \frac{1}{\sigma_p} \end{pmatrix}_{p \times p}$$

# Σχέση πίνακα συσχέτισης και πίνακα συνδιακύμανσης

$$\mathbf{P} = \mathbf{D}^{-1/2} \mathbf{\Sigma} \mathbf{D}^{-1/2}$$

**Απόδειξη** Παρατηρήστε ότι

$$\begin{aligned} \mathbf{D}^{-1/2} \mathbf{\Sigma} &= \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_p} \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{21} & \dots & \sigma_{p1} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1 & \frac{\sigma_{21}}{\sigma_1} & \dots & \frac{\sigma_{p1}}{\sigma_1} \\ \frac{\sigma_{21}}{\sigma_2} & \sigma_2 & \dots & \frac{\sigma_{p2}}{\sigma_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sigma_p} & \frac{\sigma_{p2}}{\sigma_p} & \dots & \sigma_p \end{pmatrix} \end{aligned}$$

## Σχέση πίνακα συσχέτισης και πίνακα συνδιακύμανσης

$$\begin{aligned} \mathbf{D}^{-1/2} \mathbf{\Sigma} \mathbf{D}^{-1/2} &= \begin{pmatrix} \sigma_1 & \frac{\sigma_{21}}{\sigma_1} & \cdots & \frac{\sigma_{p1}}{\sigma_1} \\ \frac{\sigma_{21}}{\sigma_2} & \sigma_2 & \cdots & \frac{\sigma_{p2}}{\sigma_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sigma_p} & \frac{\sigma_{p2}}{\sigma_p} & \cdots & \sigma_p \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_p} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \frac{\sigma_{21}}{\sigma_1 \sigma_2} & \cdots & \frac{\sigma_{p1}}{\sigma_1 \sigma_p} \\ \frac{\sigma_{21}}{\sigma_2 \sigma_1} & 1 & \cdots & \frac{\sigma_{p2}}{\sigma_2 \sigma_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sigma_p \sigma_1} & \frac{\sigma_{p2}}{\sigma_p} & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 1 & \rho_{21} & \cdots & \rho_{p1} \\ \rho_{21} & 1 & \cdots & \rho_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix} \end{aligned}$$

Επομένως,  $\mathbf{P} = \mathbf{D}^{-1/2} \mathbf{\Sigma} \mathbf{D}^{-1/2}$ . Ανάλογα αποδεικνύεται ότι  $\mathbf{\Sigma} = \mathbf{D}^{1/2} \mathbf{P} \mathbf{D}^{1/2}$ .

## Δεδομένα

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}_{n \times p} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}_{n \times p}$$

- Οι γραμμές αντιπροσωπεύουν άτομα (ερευνητικές μονάδες), ενώ οι στήλες μεταβλητές
- Οι παρατηρήσεις  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  είναι ανεξάρτητες με την ίδια κατανομή.
- Με άλλα λόγια, έχουμε ένα τυχαίο δείγμα μεγέθους  $n$  από μια  $p$ -διάστατη κατανομή.

# Αμερόληψία δειγματικού μέσου

$\bar{\mathbf{x}}_{p \times 1} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \rightarrow p$ -διάστατος δειγματικός μέσος. Το  $j$ -οστό στοιχείο είναι ο δειγματικός μέσος για την μεταβλητή  $j$ .

## Υπόθεση

Έστω  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  ένα τυχαίο δείγμα μεγέθους  $n$  από **οποιαδήποτε**  $p$ -διάστατη κατανομή με μέσο  $\boldsymbol{\mu}$  και πίνακα συνδιακύμανσης  $\boldsymbol{\Sigma}$

$$E(\bar{\mathbf{x}}) = \frac{1}{n} \sum_{i=1}^n E(\mathbf{x}_i) = \frac{n\boldsymbol{\mu}}{n} = \boldsymbol{\mu}$$

$$\text{Var}(\bar{\mathbf{x}}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(\mathbf{x}_i) = \frac{\boldsymbol{\Sigma}}{n}$$

Επομένως, ο δειγματικός μέσος  $\bar{\mathbf{x}}$  είναι αμερόληπτος εκτιμητής του πληθυσμιακού μέσου  $\boldsymbol{\mu}$  με πίνακα συνδιακύμανσης  $\frac{\boldsymbol{\Sigma}}{n}$

- $\text{Var}(\bar{\mathbf{x}}) \rightarrow \mathbf{0}$  καθώς  $n \rightarrow \infty$ .

## Δειγματικός πίνακας συνδιακύμανσης

Το  $(j, k)$  στοιχείο του  $\mathbf{S}$  είναι το  $s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$ , δηλαδή η δειγματική συνδιακύμανση των μεταβλητών  $j$  και  $k$ .

### Δειγματικός πίνακας συνδιακύμανσης

Σε μορφή πινάκων, έπεται ότι

$$\mathbf{S}_{p \times p} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

$$\mathbf{S} = \begin{pmatrix} s_1^2 & s_{21} & \dots & s_{p1} \\ s_{21} & s_2^2 & \dots & s_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{pmatrix}$$

- $s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$  και  $s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ .

# Αμεροληψία δειγματικού πίνακα συνδιακύμανσης

Παρατηρήστε ότι

$$\begin{aligned}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top &= \sum_{i=1}^n (\mathbf{x}_i \mathbf{x}_i^\top - \mathbf{x}_i \bar{\mathbf{x}}^\top - \bar{\mathbf{x}} \mathbf{x}_i^\top + \bar{\mathbf{x}} \bar{\mathbf{x}}^\top) \\&= \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top - n \bar{\mathbf{x}} \bar{\mathbf{x}}^\top - n \bar{\mathbf{x}} \bar{\mathbf{x}}^\top + n \bar{\mathbf{x}} \bar{\mathbf{x}}^\top \\&= \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top - n \bar{\mathbf{x}} \bar{\mathbf{x}}^\top\end{aligned}$$

Επιπλέον, για κάθε τυχαίο διάνυσμα  $\mathbf{x}$

$$\text{Var}(\mathbf{x}) = E(\mathbf{x} \mathbf{x}^\top) - E(\mathbf{x})E(\mathbf{x})^\top.$$

## Αμεροληψία δειγματικού πίνακα συνδιακύμανσης

$$\begin{aligned}\mathbf{S} &= \frac{1}{n-1} \left( \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top - n \overline{\mathbf{x}} \overline{\mathbf{x}}^\top \right) \Rightarrow \\ E(\mathbf{S}) &= \frac{1}{n-1} \left\{ \sum_{i=1}^n E(\mathbf{x}_i \mathbf{x}_i^\top) - n E(\overline{\mathbf{x}} \overline{\mathbf{x}}^\top) \right\} \\ &= \frac{1}{n-1} \left\{ \sum_{i=1}^n (\boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^\top) - n \left( \frac{\boldsymbol{\Sigma}}{n} + \boldsymbol{\mu} \boldsymbol{\mu}^\top \right) \right\} \\ &= \frac{1}{n-1} \left( n \boldsymbol{\Sigma} + n \boldsymbol{\mu} \boldsymbol{\mu}^\top - \boldsymbol{\Sigma} - n \boldsymbol{\mu} \boldsymbol{\mu}^\top \right) \\ &= \frac{n-1}{n-1} \boldsymbol{\Sigma} = \boldsymbol{\Sigma}.\end{aligned}$$

Ο δειγματικός πίνακας συνδιακύμανσης,  $\mathbf{S}$ , είναι αμερόληπτος εκτιμητής του  $\boldsymbol{\Sigma}$ .