Lecture Notes 1

MULTIVARIATE ANALYSIS AND PRELIMINARIES

1 Aspects of Multivariate Analysis

Reference: Johnson & Wichern (2002) *Applied Multivariate Statistical Analysis* Chapter 1.

(a) Multivariate analysis is concerned with two or more random variables.

(b) Many multivariate methods are based on an underling probability model known as the multivariate normal distribution.

(c) The objectives of the investigation for multivariate analysis include:

- Data reduction
- Sorting and grouping
- Investigation of the dependence among variables
- Prediction
- Hypothesis construction and testing

(d) The multivariate statistical methods are:

- Principal component analysis (PCA)
- Factor analysis (FA)
- Discrimination and classification analysis (DCA)
- Cluster analysis (CA)
- Multivariate linear regression model (LRM)

(e) Applications: Multivariate methods have been widely applied to many practical problems arising in

- Medicine science : (discrimination)
- Physics : (linear regression model)
- Sociology
- Business and Economics

- Environmental study
- Meteorology
- Geology
- Psychology

1.1 Multivariate Graphical Techniques

Reference: Johnson & Wichern (2002) *Applied Multivariate Statistical Analysis* Fifth Edition, Chapter 1, pages 11-49.

- Multiple Scatter Plots

- Multiple Scatter Plots
- Multiple Scatter and/or boxplots.
- Three-dimensional scatter plot for trivariate data.
- Three-dimensional scatter plots with rotation.
- Three-dimensional perspectives.
- Growth curves
- Chernoff faces
- Distances between the data points.

more than 3

2 Matrix Algebra

Reference: Johnson & Wichern (2002) *Applied Multivariate Statistical Analysis*, Chapter 2.

2.1 Vectors

An array of $\boldsymbol{x}$ of n real numbers x_1, x_2, ... , x_n is called a vector of dimension n and it is written as

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}_{n\times 1} \quad \text{or } \boldsymbol{x}^T = (x_1 \quad x_2 \quad \dots \quad x_n)_{1\times n}$$

(a) A vector can be multiply by a constant. Multiplying vector $\boldsymbol{x}$ by constant a is given by

$$a\boldsymbol{x} = \begin{pmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{pmatrix}_{n\times 1}$$

(b) Two vectors may be added. Addition of $\boldsymbol{x}$ and $\boldsymbol{y}$:

x + y

$$\boldsymbol{x} + \boldsymbol{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}_{n\times 1} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}_{n\times 1}$$

(c) The length of a vector $\mathbf{x}$ is defined by $L_x = |\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$.

(d) Multiplication of Two Vectors:

(1) Multiplication of $\boldsymbol{x}^T$ by $\boldsymbol{y}$ gives a single number. Note $L_x = \sqrt{\boldsymbol{x}^T x} = |\boldsymbol{x}|$

$$\boldsymbol{x}^T\boldsymbol{y} = (x_1 \quad x_2 \quad \dots \quad x_n)_{1\times n} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1} = x_1y_1 + x_2y_2 + \dots + x_ny_n$$

(2) Multiplication of $\boldsymbol{x}$ by $\boldsymbol{y}^T$ gives a $n \times n$ matrix.

$$\boldsymbol{x}\boldsymbol{y}^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}_{n\times 1} (y_1\ y_2\ \dots\ y_n)_{1\times n} = \begin{pmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_n \\ x_2y_1 & x_2y_2 & \dots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \dots & x_ny_n \end{pmatrix}_{n\times n}$$

2.2 Matrices

(a) A matrix is a rectangular array of real numbers. A matrix with m rows and n columns is denoted by

columns

$$\mathcal{A} = \mathcal{A}_{m\times n} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}_{m\times n} = (a_{ij})_{m\times n}$$

rows columns rows position

(b) Transpose of $\mathcal{A}$ is matrix of $n \times m$ and denoted by $\mathcal{A}^T$. That is, $\mathcal{A}^T = (a_{ji})_{n\times m}$.

(c) Multiplication of Two Matrices: $\mathcal{A}_{m\times n} = (a_{ij})_{m\times n}$ and $\mathcal{B}_{n\times p} = (b_{ij})_{n\times p}$ give matrix $\mathcal{C}_{m\times p} = (c_{ij})_{m\times p}$ where $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$.

(d) Square Matrix: If the number of rows and number of columns of a matrix are equal, then it is called a square matrix (e.g. $m = n$).

$$\mathcal{A} = \mathcal{A}_{n\times n} = \boxed{\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}_{n\times n}}$$

(e) Symmetric Matrix: If $\mathcal{A}$ is a square matrix of order n and the elements $a_{ij} = a_{ji}$ for all i and j, then $\mathcal{A}$ is called a symmetric matrix.

Note : $\mathcal{A}$ is a symmetric matrix iff $\mathcal{A}^T = \mathcal{A}$.

(f) Identity matrix: A square matrix with ones on the diagonal and zeros elsewhere is called an identity matrix. The identity matrix of order n is given by similar to 1

$$\mathcal{I}_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}_{n\times n} = \text{diag}(1\ \ 1\ \ \dots\ \ 1)$$

(g) Inverse of a Matrix: If there exists a matrix $\mathcal{B}$ for a given square matrix $\mathcal{A}$ such that $\mathcal{AB} = \mathcal{BA} = \mathcal{I}$ then $\mathcal{B}$ is called the inverse of $\mathcal{A}$ and is denoted by $\mathcal{A}^{-1}$.

(h) Orthogonal Matrix: A square matrix $\mathcal{Q}$ is said to be orthogonal if

$$\mathcal{Q}^T\mathcal{Q} = \mathcal{Q}\mathcal{Q}^T = \mathcal{I}.$$

Note : $\mathcal{Q}^{-1} = \mathcal{Q}^T$.

(i) The Trace of a Matrix: Let $\mathcal{A} = (a_{ij})_{n\times n}$, a square matrix of order n. The trace of the matrix $\mathcal{A}$ is the sum of the diagonal elements of $\mathcal{A}$ and it is denoted by $\text{tr}(\mathcal{A})$. That is,

$$\text{tr}(\mathcal{A}) = \sum_{i=1}^{n} a_{ii}.$$

(j) If $\mathcal{B}$ and $\mathcal{C}$ are square matrices, then

$$\text{tr}(\mathcal{BC}) = \text{tr}(\mathcal{CB}).$$

(k) For a quadratic form $\boldsymbol{x}^T\mathcal{A}\boldsymbol{x}$, we have

$$\boldsymbol{x}^T\mathcal{A}\boldsymbol{x} = \text{tr}(\boldsymbol{x}^T\mathcal{A}\boldsymbol{x}) = \text{tr}(\boldsymbol{x}\boldsymbol{x}^T\mathcal{A}) = \text{tr}(\mathcal{A}\boldsymbol{x}\boldsymbol{x}^T)$$

(l) Determinant: The determinant of a square matrix $\mathcal{A}$ is denoted by $|\mathcal{A}|$ or $\det(\mathcal{A})$ and it is a numerical value computed from

$$|\mathcal{A}| = \det(\mathcal{A}) = \sum_{i=1}^{n}(-1)^{i+1}a_{1i}|\mathcal{A}_{1i}|$$

if $A_{2\times2}$ $|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$ $= ad - bc$

if $A_{3\times3}$ $= \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$ $= a(ei-fh)$ $-b(di-fg)$ $+c(dh-eg)$

where $\mathcal{A}_{1i}$ is the square matrix of order $(n-1)$ obtained by deleting the first row and i^{th} column of $\mathcal{A}$.

(m) Eigenvalues and Eigenvectors: Let $\mathcal{A}$ be a square matrix of order n and $\boldsymbol{e}$ be a n dimensional vector. The values of λ and e which satisfy the equation

characteristic root

$$\mathcal{A}\boldsymbol{e} = \lambda\boldsymbol{e}$$

are respectively called the eigenvalues and eigenvectors of $\mathcal{A}$.

Note:

(1) The eigenvalues of $\mathcal{A}$ can be obtained by solving $|\mathcal{A} - \lambda\mathcal{I}| = 0$.

(2) There are n eigenvalues for a matrix of order n. λ not necessarily unique

(n) Let $(\lambda_i, \boldsymbol{e}_i)$, $i = 1, 2, ..., n$, be the eigenvalue-eigenvector pairs of $\mathcal{A}$. Then

$$\mathcal{A}\boldsymbol{e}_i = \lambda_i\boldsymbol{e}_i$$

characteristic eqn

for $i = 1, 2, ...n$. If we chose $\boldsymbol{e}_i$ such that $\boldsymbol{e}_i^T\boldsymbol{e}_i = 1$, then $\boldsymbol{e}_i$ is called the **normalized eigenvector** corresponding to λ_i.

(o) Spectral Decomposition: If $(\lambda_i, \boldsymbol{e}_i)$, $i = 1, 2, ..., n$, be the eigenvalue, normalized eigenvector pairs of the square matrix $\mathcal{A}$, then

$$\mathcal{A} = \lambda_1\boldsymbol{e}_1\boldsymbol{e}_1^T + \lambda_2\boldsymbol{e}_2\boldsymbol{e}_2^T + \ldots + \lambda_n\boldsymbol{e}_n\boldsymbol{e}_n^T.$$

This is known as the spectral decomposition of $\mathcal{A}$.

(p) The determinant of $\mathcal{A}$ can be expressed as the product of its eigenvalues. That is,

$$|\mathcal{A}| = \prod_{i=1}^{n}\lambda_i$$

(q) Positive Definite Matrix: The symmetric matrix $\mathcal{A}$ is called a positive definite matrix provided

$$\boldsymbol{x}^T \mathcal{A} \boldsymbol{x} > 0 \qquad \text{for} \quad \boldsymbol{x} \neq \mathbf{0}$$

If $\boldsymbol{x}^T \mathcal{A} \boldsymbol{x} \geq 0$ then $\mathcal{A}$ is called nonnegative (or semipositive) definite.

(r) The symmetric matrix $\mathcal{A}$ is positive definite iff the eigenvalues of $\mathcal{A}$, $\lambda_i > 0$ for all i.

(s) If $\mathcal{A}$ is nonnegative definite then $\lambda_i \geq 0$ for all i.

(t) If $\mathcal{A}$ is positive definite, then the special decomposition of the inverse

$$\mathcal{A}^{-1} = \lambda_1^{-1} \boldsymbol{e}_1 \boldsymbol{e}_1^T + \lambda_2^{-1} \boldsymbol{e}_2 \boldsymbol{e}_2^T + \ldots + \lambda_n^{-1} \boldsymbol{e}_n \boldsymbol{e}_n^T.$$

(u) Square Root of a Positive Definite Matrix: Using the special decomposition, the square root of a positive matrix is given by

$$\mathcal{A}^{\frac{1}{2}} = \lambda_1^{\frac{1}{2}} \boldsymbol{e}_1 \boldsymbol{e}_1^T + \lambda_2^{\frac{1}{2}} \boldsymbol{e}_2 \boldsymbol{e}_2^T + \ldots + \lambda_n^{\frac{1}{2}} \boldsymbol{e}_n \boldsymbol{e}_n^T.$$

3 Random Vectors

Reference: Johnson & Wichern (2002) *Applied Multivariate Statistical Analysis*, Chapter 2.

3.1 Moments of Random Vectors

A random vector is a vector whose elements are random variables. Let $\boldsymbol{X}^T = (X_1 \;\; X_2 \;\; \ldots \;\; X_p)_{1\times p}$ be a random vector. Then X_i $(i = 1, 2, \ldots p)$ are random variables.

(a) Population Mean Vector, $\boldsymbol{\mu}$ is

bold

$$\boldsymbol{\mu} = \boldsymbol{\mu}_{p\times 1} = \mathbf{E}(\boldsymbol{X}_{p\times 1}) = \begin{pmatrix} \mathbf{E}(X_1) \\ \mathbf{E}(X_2) \\ \vdots \\ \mathbf{E}(X_p) \end{pmatrix}_{p\times 1} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}_{p\times 1}$$

where $\mu_i = \mathbf{E}(X_i)$ for $i = 1, 2, \ldots, p$.

(b) Population covariance matrix, Σ is given by

$$\Sigma = \Sigma_{p\times p} = \mathbf{Cov}(\boldsymbol{X}_{p\times 1}) = \mathbf{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T].$$

This gives

covariance X_1 and X_2

$$\Sigma_{p\times p} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_{pp} \end{pmatrix}_{p\times p}$$

variance

where $\sigma_{ij} = \mathbf{Cov}(X_i, X_j) = \mathbf{E}[(X_i - \mu_i)(X_j - \mu_j)]$ and $\sigma_{ij} = \sigma_{ji}$. Note when $i = j$, $\sigma_{ii} = \sigma_i^2 = \mathbf{E}(X_i - \mu_i)^2 = \mathbf{Var}(X_i)$.

Note : Covariance matrix is a symmetric matrix, that is, $\Sigma^T = \Sigma$.

(c) Generalized Variance: The determinant of a covariance matrix is called the generalized variance, hence

$$\text{Generalized Variance} = \det(\Sigma). \qquad = |\Sigma|$$

(d) Population Correlation Matrix $\boldsymbol{\rho}$: The correlation matrix of a random vector $\boldsymbol{X}$ is given by

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{pmatrix}_{p\times p}$$

where

covariance

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} \quad \text{for} \quad i \neq j$$

variance

(e) Let $\mathcal{V} = \begin{pmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{pp} \end{pmatrix}_{p\times p} = \text{diag}(\sigma_{11} \;\; \sigma_{22} \;\; \dots \;\; \sigma_{pp})$ then, it is easily verified that

$$\boldsymbol{\rho} = \mathcal{V}^{-\frac{1}{2}}\Sigma\mathcal{V}^{-\frac{1}{2}}.$$

Note that $\mathcal{V}^{-\frac{1}{2}} = \text{diag}(1/\sqrt{\sigma_{11}} \;\; 1/\sqrt{\sigma_{22}} \;\; \dots 1/\sqrt{\sigma_{pp}})$ and $\Sigma = \mathcal{V}^{\frac{1}{2}}\boldsymbol{\rho}\mathcal{V}^{\frac{1}{2}}$

3.2 Properties of Random Vectors

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two p-variate random vectors (not necessarily normal) such that

$$\begin{aligned} \mathbf{E}(\boldsymbol{X}) &= \boldsymbol{\mu}_x, & \mathbf{E}(\boldsymbol{Y}) &= \boldsymbol{\mu}_y, \\ \mathbf{Cov}(\boldsymbol{X}) &= \Sigma_x, & \mathbf{Cov}(\boldsymbol{Y}) &= \Sigma_y, \end{aligned}$$

and $\mathbf{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \Gamma$. Consider $\boldsymbol{a}, \boldsymbol{b}$ and $\boldsymbol{c}$ are vector constant and, $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ are constant matrices. We can prove the following properties.

(a) $\mathbf{E}(\boldsymbol{a}^T\boldsymbol{X}) = \boldsymbol{a}^T\boldsymbol{\mu}_x$

(b) $\mathbf{E}(\mathcal{A}^T\boldsymbol{X} + \boldsymbol{c}) = \mathcal{A}^T\boldsymbol{\mu}_x + \boldsymbol{c}$

(c) $\mathbf{Cov}(\boldsymbol{a}^T\boldsymbol{X}, \boldsymbol{b}^T\boldsymbol{X}) = \boldsymbol{a}^T\Sigma_x\boldsymbol{b} = \boldsymbol{b}^T\Sigma_x\boldsymbol{a}$

(d) $\mathbf{Cov}(\boldsymbol{a}^T\boldsymbol{X}) = \boldsymbol{a}^T\Sigma_x\boldsymbol{a}$

(e) $\mathbf{Cov}(\mathcal{D}\boldsymbol{X}) = \mathcal{D}\Sigma_x\mathcal{D}^T$

(f) $\mathbf{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathbf{E}\{(\boldsymbol{X} - \mathbf{E}(\boldsymbol{X}))(\boldsymbol{Y} - \mathbf{E}(\boldsymbol{Y})^T\} = \mathbf{Cov}(\boldsymbol{Y}, \boldsymbol{X})^T$

(g) $\mathbf{Cov}(\mathcal{A}^T\boldsymbol{X}, \mathcal{B}^T\boldsymbol{X}) = \mathcal{A}^T\Sigma_x\mathcal{B}$

(h) $\mathbf{Cov}(\boldsymbol{a}^T\boldsymbol{X}, \boldsymbol{b}^T\boldsymbol{Y}) = \boldsymbol{a}^T\mathbf{Cov}(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{b}$

(i) $\mathbf{Cov}(\mathcal{A}^T\boldsymbol{X}, \mathcal{B}^T\boldsymbol{Y}) = \mathcal{A}^T\mathbf{Cov}(\boldsymbol{X}, \boldsymbol{Y})\mathcal{B}$

3.3 Linear combinations of random variables

Consider random variables X_1 and X_2 and a and b are constants, then

$$\mathbf{E}(aX_1 + bX_2) = a\mathbf{E}(X_1) + b\mathbf{E}(X_2) = a\mu_1 + b\mu_2$$

using additional properties, covariance of aX_1 and bX_2 is

$$\mathbf{Cov}(aX_1, bX_2) = \mathbf{E}[(aX_1 - a\mu_1)(bX_2 - b\mu_2)] = ab\mathbf{Cov}(X_1, X_2) = ab\sigma_{12}$$

and

$$\begin{aligned}\mathbf{Var}(aX_1 + bX_2) &= a^2\mathbf{Var}(X_1) + b^2\mathbf{Var}(X_2) + 2ab\mathbf{Cov}(X_1, X_2)\\ &= a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12} = a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\sigma_1\sigma_2\rho_{12}\end{aligned}$$

(a) If X_1 and X_2 are independent random variable, then $\rho_{12} = 0$, hence

$$\mathbf{Var}(aX_1 + bX_2) = a^2\sigma_1^2 + b^2\sigma_2^2$$

(b) Let $\mathbf{c}^T = (a \;\; b)$, $\boldsymbol{\mu}^T = (\mu_1 \;\; \mu_2)$ and $\boldsymbol{X}^T = (X_1 \;\; X_2)$ then

$$\mathbf{E}(\mathbf{c}^T\boldsymbol{X}) = \mathbf{E}(ab)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = (ab)\begin{pmatrix} \mathbf{E}(X_1) \\ \mathbf{E}(X_2) \end{pmatrix} = \mathbf{E}(ab)\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \mathbf{c}^T\boldsymbol{\mu}$$

Note that $\mathbf{E}(\mathbf{c}^T\boldsymbol{X}) = \mathbf{E}(aX_1 + bX_2) = a\mu_1 + b\mu_2 = \mathbf{c}^T\boldsymbol{\mu}$.

(c) Let

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

then

$$\mathbf{Var}(\mathbf{c}^T\boldsymbol{X}) = (a \;\; b)\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{c}^T\Sigma\mathbf{c}$$

Note that $\mathbf{Var}(\mathbf{c}^T\boldsymbol{X}) = \mathbf{Var}(aX_1 + bX_2) = a^2\sigma_{11} + +2ab\sigma_{12} + b^2\sigma_{22} = \mathbf{c}^T\Sigma\mathbf{c}$.

(d) The linear combination $\mathbf{c}^T\boldsymbol{X} = c_1X_1 + c_2X_2 + \cdots + c_pX_p$ has

$$\begin{aligned}&\text{Mean} &&= \mathbf{E}(\mathbf{c}^T\boldsymbol{X}) = \mathbf{c}^T\boldsymbol{\mu} \qquad \text{and}\\ &\text{Variance} &&= \mathbf{Var}(\mathbf{c}^T\boldsymbol{X}) = \mathbf{c}^T\Sigma\mathbf{c}\end{aligned}$$

where $\boldsymbol{\mu} = \mathbf{E}(\boldsymbol{X})$ and $\Sigma = \mathbf{Cov}(\boldsymbol{X})$.

(e) In general consider the q linear combinations of the p random variables $X_1, X_2, \ldots, X_p$.

$$\begin{aligned}Z_1 &= a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p\\ Z_2 &= a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p\\ \vdots & \quad \vdots\\ Z_q &= a_{q1}X_1 + a_{q2}X_2 + \ldots + a_{qp}X_p\end{aligned}$$

This linear combinations can be rewritten as $\boldsymbol{Z}_{q\times 1} = \mathcal{A}_{q\times p}\boldsymbol{X}_{p\times 1}$, therefore we obtain

$$\boldsymbol{\mu}_z = \mathbf{E}(\boldsymbol{Z}) = \mathbf{E}(\mathcal{A}\boldsymbol{X}) = \mathcal{A}\boldsymbol{\mu}_x$$

and

$$\Sigma_z = \mathbf{Cov}(\boldsymbol{Z}) = \mathbf{Cov}(\mathcal{A}\boldsymbol{X}) = \mathcal{A}\Sigma_x\mathcal{A}^T$$

where $\boldsymbol{\mu}_x$ and Σ_x are respectively the mean vector and covariance matrix of $\boldsymbol{X}$.