TERMINOLOGY: Suppose that $Y_1, Y_2, ..., Y_n$ are random variables. We call

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

a **random vector**. The joint pdf of $\mathbf{Y}$ is denoted by $f_{\mathbf{Y}}(\mathbf{y})$.

DEFINTION: Suppose that $E(Y_i) = \mu_i$, $\text{var}(Y_i) = \sigma_i^2$, for $i = 1, 2, ..., n$, and $\text{cov}(Y_i, Y_j) = \sigma_{ij}$, for $i \neq j$. The **mean** of $\mathbf{Y}$ is

$$\boldsymbol{\mu} = E(\mathbf{Y}) = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}.$$

The **variance-covariance matrix** of $\mathbf{Y}$ is

$$\boldsymbol{\Sigma} = \text{cov}(\mathbf{Y}) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}.$$

NOTE: Note that $\boldsymbol{\Sigma}$ contains the variances $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$ on the diagonal and the $\binom{n}{2}$ covariance terms $\text{cov}(Y_i, Y_j)$, for $i < j$, as the elements strictly above the diagonal. Since $\text{cov}(Y_i, Y_j) = \text{cov}(Y_j, Y_i)$, it follows that $\boldsymbol{\Sigma}$ is symmetric.

EXAMPLE: Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample with mean $E(Y_i) = \mu$ and variance $\text{var}(Y_i) = \sigma^2$ and let $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$. Then $\boldsymbol{\mu} = E(\mathbf{Y}) = \mu\mathbf{1}_n$ and $\boldsymbol{\Sigma} = \text{cov}(\mathbf{Y}) = \sigma^2\mathbf{I}_n$.

EXAMPLE: Consider the GM linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. In this model, the random errors $\epsilon_1, \epsilon_2, ..., \epsilon_n$ are uncorrelated random variables with zero mean and constant variance σ^2. We have $E(\boldsymbol{\epsilon}) = \mathbf{0}_{n\times 1}$ and $\text{cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_n$.

TERMINOLOGY: Suppose that $Z_{11}, Z_{12}, ..., Z_{np}$ are random variables. We call

$$\mathbf{Z}_{n\times p} = \begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1p} \\ Z_{21} & Z_{22} & \cdots & Z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{np} \end{pmatrix}$$

a **random matrix**. The mean of $\mathbf{Z}$ is

$$E(\mathbf{Z}) = \begin{pmatrix} E(Z_{11}) & E(Z_{12}) & \cdots & E(Z_{1p}) \\ E(Z_{21}) & E(Z_{22}) & \cdots & E(Z_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ E(Z_{n1}) & E(Z_{n2}) & \cdots & E(Z_{np}) \end{pmatrix}_{n\times p}.$$

Result RV1. Suppose that $\mathbf{Y}$ is a random vector with mean $\boldsymbol{\mu}$. Then

$$\boldsymbol{\Sigma} = \text{cov}(\mathbf{Y}) = E[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'] = E(\mathbf{Y}\mathbf{Y}') - \boldsymbol{\mu}\boldsymbol{\mu}'.$$

Proof. That $\text{cov}(\mathbf{Y}) = E[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})']$ follows straightforwardly from the definition of variance and covariance in the scalar case. Showing this equals $E(\mathbf{Y}\mathbf{Y}') - \boldsymbol{\mu}\boldsymbol{\mu}'$ is simple algebra. □

DEFINITION: Suppose that $\mathbf{Y}_{p\times 1}$ and $\mathbf{X}_{q\times 1}$ are random vectors with means $\boldsymbol{\mu}_{\mathbf{Y}}$ and $\boldsymbol{\mu}_{\mathbf{X}}$, respectively. The covariance between $\mathbf{Y}$ and $\mathbf{X}$ is the $p \times q$ matrix defined by

$$\text{cov}(\mathbf{Y}, \mathbf{X}) = E\{(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})'\} = (\sigma_{ij})_{p\times q},$$

where

$$\sigma_{ij} = E[\{Y_i - E(Y_i)\}\{X_j - E(X_j)\}] = \text{cov}(Y_i, X_j).$$

DEFINITION: Random vectors $\mathbf{Y}_{p\times 1}$ and $\mathbf{X}_{q\times 1}$ are **uncorrelated** if $\text{cov}(\mathbf{Y}, \mathbf{X}) = \mathbf{0}_{p\times q}$.

Result RV2. If $\text{cov}(\mathbf{Y}, \mathbf{X}) = \mathbf{0}$, then $\text{cov}(\mathbf{Y}, \mathbf{a} + \mathbf{B}\mathbf{X}) = \mathbf{0}$, for all nonrandom conformable $\mathbf{a}$ and $\mathbf{B}$. That is, $\mathbf{Y}$ is uncorrelated with any linear function of $\mathbf{X}$.

TERMINOLOGY: Suppose that $\text{var}(Y_i) = \sigma_i^2$, for $i = 1, 2, ..., n$, and $\text{cov}(Y_i, Y_j) = \sigma_{ij}$, for $i \neq j$. The **correlation matrix** of $\mathbf{Y}$ is the $n \times n$ matrix

$$\mathbf{R} = (\rho_{ij}) = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{pmatrix},$$

where, recall, the correlation ρ_{ij} is given by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j},$$

for $i, j = 1, 2, ..., n$.

TERMINOLOGY: Suppose that $Y_1, Y_2, ..., Y_n$ are random variables and that $a_1, a_2, ..., a_n$ are constants. Define $\mathbf{a} = (a_1, a_2, ..., a_n)'$ and $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$. The random variable

$$X = \mathbf{a}'\mathbf{Y} = \sum_{i=1}^{n} a_i Y_i$$

is called a **linear combination** of $Y_1, Y_2, ..., Y_n$.

Result RV3. If $\mathbf{a} = (a_1, a_2, ..., a_n)'$ is a vector of constants and $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ is a random vector with mean $\boldsymbol{\mu} = E(\mathbf{Y})$, then

$$E(\mathbf{a}'\mathbf{Y}) = \mathbf{a}'\boldsymbol{\mu}.$$

Proof. The quantity $\mathbf{a}'\mathbf{Y}$ is a scalar so $E(\mathbf{a}'\mathbf{Y})$ is also a scalar. Note that

$$E(\mathbf{a}'\mathbf{Y}) = E\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n} a_i E(Y_i) = \sum_{i=1}^{n} a_i \mu_i = \mathbf{a}'\boldsymbol{\mu}. \ \square$$

Result RV4. Suppose that $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ is a random vector with mean $\boldsymbol{\mu} = E(\mathbf{Y})$, let $\mathbf{Z}$ be a random matrix, and let $\mathbf{A}$ and $\mathbf{B}$ ($\mathbf{a}$ and $\mathbf{b}$) be nonrandom conformable matrices (vectors). Then

1. $E(\mathbf{AY}) = \mathbf{A}\mu$
2. $E(\mathbf{a}'\mathbf{Zb}) = \mathbf{a}'E(\mathbf{Z})\mathbf{b}$.
3. $E(\mathbf{AZB}) = \mathbf{A}E(\mathbf{Z})\mathbf{B}$.

Result RV5. If $\mathbf{a} = (a_1, a_2, ..., a_n)'$ is a vector of constants and $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ is a random vector with mean $\boldsymbol{\mu} = E(\mathbf{Y})$ and covariance matrix $\boldsymbol{\Sigma} = \text{cov}(\mathbf{Y})$, then

$$\text{var}(\mathbf{a}'\mathbf{Y}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}.$$

Proof. The quantity $\mathbf{a}'\mathbf{Y}$ is a scalar random variable, and its variance is given by

$$\text{var}(\mathbf{a}'\mathbf{y}) = E\{(\mathbf{a}'\mathbf{Y} - \mathbf{a}'\boldsymbol{\mu})^2\} = E[\{\mathbf{a}'(\mathbf{Y} - \boldsymbol{\mu})\}^2] = E\{\mathbf{a}'(\mathbf{Y} - \boldsymbol{\mu})\mathbf{a}'(\mathbf{Y} - \boldsymbol{\mu})\}.$$

But, note that $\mathbf{a}'(\mathbf{Y} - \boldsymbol{\mu})$ is a scalar, and hence equals $(\mathbf{Y} - \boldsymbol{\mu})'\mathbf{a}$. Using this fact, we can rewrite the last expectation to get

$$E\{\mathbf{a}'(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'\mathbf{a}\} = \mathbf{a}'E\{(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'\}\mathbf{a} = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}. \ \square$$

Result RV6. Suppose that $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ is a random vector with covariance matrix $\boldsymbol{\Sigma} = \text{cov}(\mathbf{Y})$, and let $\mathbf{a}$ and $\mathbf{b}$ be conformable vectors of constants. Then

$$\text{cov}(\mathbf{a}'\mathbf{Y}, \mathbf{b}'\mathbf{Y}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{b}.$$

Result RV7. Suppose that $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ is a random vector with mean $\boldsymbol{\mu} = E(\mathbf{Y})$ and covariance matrix $\boldsymbol{\Sigma} = \text{cov}(\mathbf{Y})$. Let $\mathbf{b}$, $\mathbf{A}$, and $\mathbf{B}$ denote nonrandom conformable vectors/matrices. Then

1. $E(\mathbf{AY} + \mathbf{b}) = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$
2. $\text{cov}(\mathbf{AY} + \mathbf{b}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$
3. $\text{cov}(\mathbf{AY}, \mathbf{BY}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}'$.

Result RV8. A variance-covariance matrix $\mathbf{\Sigma} = \text{cov}(\mathbf{Y})$ is nonnegative definite.
Proof. Suppose that $\mathbf{Y}_{n\times 1}$ has variance-covariance matrix $\mathbf{\Sigma}$. We need to show that $\mathbf{a}'\mathbf{\Sigma}\mathbf{a} \geq 0$, for all $\mathbf{a} \in \mathcal{R}^n$. Consider $X = \mathbf{a}'\mathbf{Y}$, where $\mathbf{a}$ is a conformable vector of constants. Then, X is scalar and $\text{var}(X) \geq 0$. But, $\text{var}(X) = \text{var}(\mathbf{a}'\mathbf{Y}) = \mathbf{a}'\mathbf{\Sigma}\mathbf{a}$. Since $\mathbf{a}$ is arbitrary, the result follows. $\square$

Result RV9. If $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ is a random vector with mean $\boldsymbol{\mu} = E(\mathbf{Y})$ and covariance matrix $\mathbf{\Sigma}$, then $P\{(\mathbf{Y} - \boldsymbol{\mu}) \in \mathcal{C}(\mathbf{\Sigma})\} = 1$.
Proof. Without loss, take $\boldsymbol{\mu} = \mathbf{0}$, and let $\mathbf{M}_{\mathbf{\Sigma}}$ be the perpendicular projection matrix onto $\mathcal{C}(\mathbf{\Sigma})$. We know that $\mathbf{Y} = \mathbf{M}_{\mathbf{\Sigma}}\mathbf{Y} + (\mathbf{I} - \mathbf{M}_{\mathbf{\Sigma}})\mathbf{Y}$ and that

$$E\{(\mathbf{I} - \mathbf{M}_{\mathbf{\Sigma}})\mathbf{Y}\} = (\mathbf{I} - \mathbf{M}_{\mathbf{\Sigma}})E(\mathbf{Y}) = \mathbf{0},$$

since $\boldsymbol{\mu} = E(\mathbf{Y}) = \mathbf{0}$. Also,

$$\text{cov}\{(\mathbf{I} - \mathbf{M}_{\mathbf{\Sigma}})\mathbf{Y}\} = (\mathbf{I} - \mathbf{M}_{\mathbf{\Sigma}})\mathbf{\Sigma}(\mathbf{I} - \mathbf{M}_{\mathbf{\Sigma}})' = (\mathbf{\Sigma} - \mathbf{M}_{\mathbf{\Sigma}}\mathbf{\Sigma})(\mathbf{I} - \mathbf{M}_{\mathbf{\Sigma}})' = \mathbf{0},$$

since $\mathbf{M}_{\mathbf{\Sigma}}\mathbf{\Sigma} = \mathbf{\Sigma}$. Thus, we have shown that $P\{(\mathbf{I} - \mathbf{M}_{\mathbf{\Sigma}})\mathbf{Y} = \mathbf{0}\} = 1$, which implies that $P(\mathbf{Y} = \mathbf{M}_{\mathbf{\Sigma}}\mathbf{Y}) = 1$. Since $\mathbf{M}_{\mathbf{\Sigma}}\mathbf{Y} \in \mathcal{C}(\mathbf{\Sigma})$, we are done. $\square$

IMPLICATION: Result RV9 says that there exists a subset $\mathcal{C}(\mathbf{\Sigma}) \subseteq \mathcal{R}^n$ that contains $\mathbf{Y}$ with probability one (i.e., almost surely). If $\mathbf{\Sigma}$ is positive semidefinite (psd), then $\mathbf{\Sigma}$ is singular and $\mathcal{C}(\mathbf{\Sigma})$ is concentrated in a subspace of $\mathcal{R}^n$, where the subspace has dimension $r = r(\mathbf{\Sigma})$, $r < n$. In this situation, the pdf of $\mathbf{Y}$ may not exist.

Result RV10. Suppose that $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are $n \times 1$ vectors and that $\mathbf{X} = \mathbf{Y} + \mathbf{Z}$. Then

1. $E(\mathbf{X}) = E(\mathbf{Y}) + E(\mathbf{Z})$
2. $\text{cov}(\mathbf{X}) = \text{cov}(\mathbf{Y}) + \text{cov}(\mathbf{Z}) + 2\text{cov}(\mathbf{Y}, \mathbf{Z})$
3. if $\mathbf{Y}$ and $\mathbf{Z}$ are uncorrelated, then $\text{cov}(\mathbf{X}) = \text{cov}(\mathbf{Y}) + \text{cov}(\mathbf{Z})$.