2020

Theory of Principal Components for Applications in Exploratory Crime Analysis and Clustering

Daniel Silva
Minnesota State University, Mankato

Theory of Principal Components for Applications in Exploratory Crime Analysis and Clustering

By

Daniel Silva

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

In

Applied Statistics

Minnesota State University, Mankato

Mankato, Minnesota

May 2020

Date: April 10th 2020

Title: Theory of Principal Components for Applications in Exploratory Crime Analysis and Clustering

Student's Name: Daniel Silva

This thesis has been examined and approved by the following members of the student's committee.

____

Advisor/Chairperson,

Dr Galkande (Iresha) Premarathna

Professor of Statistics,

Minnesota State University, Mankato.

____

Committee Member,

Dr Mezbahur Rahman,

Professor of Statistics,

Minnesota State University, Mankato.

____

Committee Member,

Dr Deepak Sanjel,

Professor of Statistics,

Minnesota State University, Mankato.

Abstract

Theory of Principal Components for Applications

in Exploratory Crime Analysis and Clustering

By Daniel Silva

Master of Science in Applied Statistics

Minnesota State University, Mankato

Mankato, Minnesota, 2020

The purpose of this paper is to develop the theory of principal components analysis succinctly from the fundamentals of matrix algebra and multivariate statistics. Principal components analysis is sometimes used as a descriptive technique to explain the variance-covariance or correlation structure of a dataset. However, most often, it is used as a dimensionality reduction technique to visualize a high dimensional dataset in a lower dimensional space. Principal components analysis accomplishes this by using the first few principal components, provided that they account for a substantial proportion of variation in the original dataset. In the same way, the first few principal components can be used as inputs into a cluster analysis in order to combat the curse of dimensionality and optimize the runtime for large datasets. The application portion of this paper will apply these methods to a US Crime 2018 dataset extracted from the Uniform Crime Reports on the FBI's website.

Table of Contents

Chapter 1

Introduction

Principal components analysis (PCA) is multivariate statistical method that seeks to transform a set of correlated variables $X_1, X_2, \dots, X_p$ into a new set of uncorrelated variables $Y_1, Y_2, \dots, Y_p$ that retain the total system variation. These new variables are called the principal components. Each principal component $Y_1, Y_2, \dots, Y_p$ is a distinct linear combination of the original variables $X_1, X_2, \dots, X_p$ derived in decreasing order of importance in the sense that Y_1 accounts for as much of the variation in the original system amongst all other linear combinations $Y_2, \dots, Y_p$. Then Y_2 is chosen to account for as much as possible of the remaining system variation, subject to being uncorrelated with Y_1. Analogously, Y_i is chosen to account for as much as possible of the remaining system variation, subject to being uncorrelated with $Y_1, Y_2, \dots, Y_{i-1}$.

The general hope of PCA is that the first few components will account for a substantial proportion of the variation in the original system, $X_1, X_2, \dots, X_p$, and can, consequently, be used to provide a lower-dimensional summary of these variables [1, p. 41]. These first few principal components may then replace the original $X_1, X_2, \dots, X_p$ and can be used for descriptives, graphical interpretations, and even inputs into another analysis, with minimal loss of information. That is why principal components analysis is often thought of as a dimensionality reduction technique as well as an interpretive aid in explained the original variables.

1.1 Theory Structure

In order to get a proper treatment of PCA, one needs a couple preliminaries including matrix algebra, multivariate population theory, and multivariate sample theory.

Matrix algebra is the backbone of multivariate statistics. Chapter 2 devotes itself to covering all essential notations and concepts necessary to understand later chapters. This includes, but is not limited to, vector/matrix notations, inner-product, matrix multiplication, independence, square matrices, orthogonal matrices, eigenvalues and eigenvectors, and matrix maximization of quadratic forms.

Covering matrix algebra before multivariate population theory is critical because it bridges the gap from one's knowledge of univariate population theory to multivariate population theory. Chapter 3, Multivariate Population Theory, covers population random matrices, random vectors, mean vectors, variance-covariance and correlations matrices, and the corresponding theory related to linear combination used directly in the treatment of population PCA. Further, the same topics, as above, are extended to standardized multivariate populations.

Chapter 4, Multivariate Sample Theory, follows directly from Chapter 3. It is paramount in understanding how one goes from population principal components to sample principal components. New concepts of multivariate random samples will be derived from concepts of matrix algebra, multivariate population theory, and univariate random samples learned in one's previous coursework. Then, the sample equivalents to Chapter 3 will be covered; including those related to standardized multivariate populations and linear combinations.

Chapter 5 is devoted to the main topic of PCA. Here we will cover population principal components for unstandardized and standardized continuous random variables. Similarly, we will cover sample principal components for unstandardized and standardized multivariate random samples.

1.2 Application Background and Structure

Local law enforcement agency across the United States collect data on violent and property crimes. Every year, the FBI compiles, publishes, and archives this data in the Uniform Crime Reports (UCR). The UCR Program's primary objective is to generate reliable information for use in law enforcement administration, operation, management, and analytics.

Violent crime definitions according to the FBI are:

- Murder and nonnegligent manslaughter: the willful (nonnegligent) killing of one human being by another.
- Rape: The penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim.

- Robbery: The taking or attempting to take anything of value from the care, custody, or control of a person or persons by force or threat of force or violence and/or by putting the victim in fear.
- Aggravated assault: An unlawful attack by one person upon another for the purpose of inflicting severe or aggravated bodily injury. This type of assault usually is accompanied by the use of a weapon or by means likely to produce death or great bodily harm. Simple assaults are excluded.

Property crime definitions according to the FBI are:

- Burglary (breaking or entering): The unlawful entry of a structure to commit a felony or a theft. Attempted forcible entry is included.
- Larceny-theft (except motor vehicle theft): The unlawful taking, carrying, leading, or riding away of property from the possession or constructive possession of another. Examples are thefts of bicycles, motor vehicle parts and accessories, shoplifting, pocket-picking, or the stealing of any property or article that is not taken by force and violence or by fraud. Attempted larcenies are included. Embezzlement, confidence games, forgery, check fraud, etc., are excluded.
- Motor vehicle theft: The theft or attempted theft of a motor vehicle. A motor vehicle is self-propelled and runs on land surface and not on rails. Motorboats, construction equipment, airplanes, and farming equipment are specifically excluded from this category [2].

For our application, Chapter 6, we shall use the UCR's US Crime 2018 data for metropolitan statistical areas. Where a metropolitan statistical area is defined by a city with surrounding suburbs that are connected by some economic factors. One disclaimer is our analysis is not meant to rank local or federal law enforcement agencies based on the crime rates in their respective regions. Our analysis is only meant to group metropolitan statistical areas with similar crime profiles and compare their group averages to each-other and to the national averages. Also, note that crimes are generally underreported.

The first step in our analysis will be of a univariate nature. We will calculate descriptives and assess the shape of each of the seven crime distributions. For example, checking whether the parent distribution is perhaps normal or even lognormal. In addition, we will look at the tail-ends of the distributions checking for univariate outliers. The second step is a bivariate distribution analysis. We will graphically visualize the correlation matrix. In addition, we will look at contour- and scatter- plots of the pairs of variables. The third step will be a short multivariate distribution analysis where we will solely test for multivariate normality.

Next, we will standardize the US Crime 2018 data to prepare it for PCA. It is common practice to do this when the ranges of the variables are largely different. Once this is done, we can calculate the sample principal components. Topics of interest are explained variance by sample principal component and contributions of standardized variables to each sample principal component. Also, one can attempt to interpret the sample principal component dimensions in the context of the subject matter--crime. Then, one can look at correlations of standardized variables with the sample principal components. Finally, one can create scatterplots of the first few sample principal components and look for clusters of metropolitan statistical areas or potential multivariate outliers.

After this we will use cluster analysis to attempt to meaningfully group (or profile) metropolitan statistical areas with similar crime attributes. We will use two sets of inputs (1) the Standardized Crime 2018 variables and (2) the first three sample principal components. Three cluster algorithms will be used k-Means, Ward's method, and Average method with both sets of inputs. This will leave use with six cluster assignments to compare and contrast graphically and via their respective cluster mean vectors.

Chapter 2

Matrix Algebra

2.1 Vectors

Definition 2.1.1 (Vector)**.** *A* $n \times 1$ *dimensional array* $\underset{(n\times 1)}{\mathbf{x}}$ of n *real numbers* $x_1, x_2, \dots, x_j, \dots, x_n$ $(n - tuple)$ *is called a* **vector**, *and in general, is denoted by a boldfaced, lowercase letter. It is written as*

$$\underset{(n\times 1)}{\mathbf{x}} = \underset{(n\times 1)}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}}$$

[3, pp. 49, 82]*.*

A vector $\underset{(n\times 1)}{\mathbf{x}}$ can be represented geometrically as a directed line in n dimensions with component x_1 along the 1th axis, x_2 along the 2nd axis,…, x_j along the jth axis,…, and x_n along the nth axis [3, p. 50].

Definition 2.1.2 (Vector Transpose). *A* $1\times n$ *dimensional array* $\underset{(1\times n)}{\mathbf{x}'}$ *of* n *real numbers* $x_1, x_2, \dots, x_j, \dots, x_n$ $(n - tuple)$ *is called a* **vector transpose**. *It is written as*

$$\underset{(1\times n)}{\mathbf{x}'} = \underset{(1\times n)}{[x_1, x_2, \dots, x_j, \dots, x_n]}$$

where the prime denotes the operation of transposing a column $\underset{(n\times 1)}{\mathbf{x}}$ *to a row* $\underset{(1\times n)}{\mathbf{x}'}$ [3, p. 49].

Definition 2.1.3 (Zero-Vector). $\underset{(n\times 1)}{\mathbf{0}}$ **vector** *is a* $n \times 1$ *dimensional array of* $0's$. *It is written as*

$$\underset{(n\times 1)}{\mathbf{0}} = \underset{(n\times 1)}{\begin{bmatrix} 0_1 \\ 0_2 \\ \vdots \\ 0_j \\ \vdots \\ 0_n \end{bmatrix}}$$

often thought of as the origin in $n -$ *space*.

Definition 2.1.4 (One Vector). $\underset{(n\times 1)}{\mathbf{1}}$ **vector** *is a* $n \times 1$ *dimensional array of* $1's$. *It is written as*

$$\underset{(n\times 1)}{\mathbf{1}} = \underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_j \\ \vdots \\ 1_n \end{bmatrix}}$$

Definition 2.1.5 (Scalar Multiplication). *Let c be an arbitrary scalar. Then the product* $\underset{(n\times 1)}{c\mathbf{x}}$ *is a vector with jth entry cx_j. It is written as*

$$\underset{(n\times 1)}{c\mathbf{x}} = c \cdot \underset{(n\times 1)}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}} = \underset{(n\times 1)}{\begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_j \\ \vdots \\ cx_n \end{bmatrix}}$$

[3, pp. 50, 82].

Definition 2.1.6 (Vector Addition). *The sum of two vectors* $\underset{(n\times 1)}{\mathbf{x}}$ *and* $\underset{(n\times 1)}{\mathbf{y}}$, *each having the same number of entries, is the vector*

$$\underset{(n\times 1)}{\mathbf{z}} = \underset{(n\times 1)}{\mathbf{x}} + \underset{(n\times 1)}{\mathbf{y}} \text{ with } j\text{th entry } z_j = x_j + y_j$$

That is,

$$\underset{(n\times 1)}{\mathbf{z}} = \underset{(n\times 1)}{\mathbf{x}} + \underset{(n\times 1)}{\mathbf{y}} = \underset{(n\times 1)}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}} + \underset{(n\times 1)}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \\ \vdots \\ y_n \end{bmatrix}} = \underset{(n\times 1)}{\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_j + y_j \\ \vdots \\ x_n + y_n \end{bmatrix}} = \underset{(n\times 1)}{\mathbf{x} + \mathbf{y}}$$

[3, pp. 51, 83].

The sum of two vectors emanating from the origin $\underset{(n\times 1)}{\mathbf{0}}$ is the diagonal of the parallelogram formed with the two original vectors as adjacent sides [3, p. 51].

Definition 2.1.7 (Vector Space). *The space of all real* $n-tuples$ *(vectors), with scalar multiplication and vector addition, is called a* **vector space** [3, p. 83].

Definition 2.1.8 (Linear Span). *The vector*

$$\underset{(n\times1)}{\mathbf{y}} = a_1\underset{(n\times1)}{\mathbf{x}_1} + a_2\underset{(n\times1)}{\mathbf{x}_2} + \cdots + a_k\underset{(n\times1)}{\mathbf{x}_k} + \cdots + a_p\underset{(n\times1)}{\mathbf{x}_p}$$

is a linear combination of the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_p$ *in* $\mathbb{R}^n$ *where* $a_1, a_2, \dots, a_k, \dots, a_p$ *are real. The set of all linear combinations of* $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_p$ *is called their* **linear span***, denoted,* $\operatorname{span}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_p)$ [3, p. 83], [4, p. 114].

Definition 2.1.9 (Linearly Dependent). *A set of vectors* $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_p$ *is said to be* **linearly dependent** *if there exist* p *numbers* $(a_1, a_2, \dots, a_k, \dots, a_p)$*, not all zero, such that*

$$a_1\underset{(n\times1)}{\mathbf{x}_1} + a_2\underset{(n\times1)}{\mathbf{x}_2} + \cdots + a_k\underset{(n\times1)}{\mathbf{x}_k} + \cdots + a_p\underset{(n\times1)}{\mathbf{x}_p} = \underset{(n\times1)}{\mathbf{0}}$$

[3, p. 83].

If one of the vectors, for example, $\underset{(n\times1)}{\mathbf{x}_k}$, is $\underset{(n\times1)}{\mathbf{0}}$, the set is linearly dependent (Let a_k be the only nonzero coefficient). Linear dependence implies that at least one vector in the set can be written as a linear combination of the other vectors. Vectors of the same dimension that are not linearly dependent are said to be **linearly independent** [3, pp. 53, 83].

Definition 2.1.10 (Basis). *Any set of* n *linearly independent vectors is called a* **basis** *for the vector space of all* $n-$ *tuples of real numbers* [3, p. 84].

Result 2.1.1. *Every vector can be expressed as a unique linear combination of a fixed basis* [3, p. 84].

Definition 2.1.11 (Inner Product). *The* **inner** *(or dot)* **product** *of two vectors* $\underset{(n\times1)}{\mathbf{x}}$ *and* $\underset{(n\times1)}{\mathbf{y}}$ *with the same number of entries is defined as the sum of component products:*

$$\underset{(1\times n)}{\mathbf{x}'} \cdot \underset{(n\times1)}{\mathbf{y}} = \underset{(1\times n)}{[x_1, x_2, \dots, x_j, \dots, x_n]} \cdot \underset{(n\times1)}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \\ \vdots \\ y_n \end{bmatrix}} = x_1y_1 + x_2y_2 + \cdots + x_jy_j + \cdots + x_ny_n$$

or

$$\underset{(1\times n)}{\mathbf{y}'} \cdot \underset{(n\times1)}{\mathbf{x}} = \underset{(1\times n)}{[y_1, y_2, \dots, y_j, \dots, y_n]} \cdot \underset{(n\times1)}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}} = y_1x_1 + y_2x_2 + \cdots + y_jx_j + \cdots + y_nx_n$$

[3, pp. 52, 85].

Hence,

$$\underset{(1\times n)}{\mathbf{x}'} \cdot \underset{(n\times1)}{\mathbf{y}} = \underset{(1\times n)}{\mathbf{y}'} \cdot \underset{(n\times1)}{\mathbf{x}}.$$

Definition 2.1.12 (Length of a Vector)**.** *A vector has both direction and length. The* **length of a vector** $\underset{(n\times 1)}{\mathbf{x}}$ *of* n *elements emanating from the origin* $\underset{(n\times 1)}{\mathbf{0}}$ *is given by the Pythagorean formula:*

$$\begin{aligned}
&\textit{length of } \underset{(n\times 1)}{\mathbf{x}} \\
&= \underset{(1\times 1)}{L_{\mathbf{x}}} \\
&= \sqrt{\underset{(1\times n)}{\mathbf{x}'} \cdot \underset{(n\times 1)}{\mathbf{x}}} \\
&= \sqrt{\underset{(1\times n)}{\left[x_1, x_2, \dots, x_j, \dots, x_n\right]} \cdot \underset{(n\times 1)}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}}} \\
&= \sqrt{x_1^2 + x_2^2 + \cdots + x_j^2 + \cdots + x_n^2}
\end{aligned}$$

[3, p. 84]*.*

Multiplication by c does not change the direction of the vector $\underset{(n\times 1)}{\mathbf{x}}$ if $c > 0$. However, a negative value of c creates a vector with a direction opposite that of $\underset{(n\times 1)}{\mathbf{x}}$. From $L_{c\mathbf{x}} = |c|L_{\mathbf{x}}$ it is clear that $\underset{(n\times 1)}{\mathbf{x}}$ is expanded if $|c| > 1$ and contracted if $0 < |c| < 1$. Choosing $c = L_{\mathbf{x}}^{-1}$, we obtain the *unit vector* $L_{\mathbf{x}}^{-1}\underset{(n\times 1)}{\mathbf{x}}$, which has length 1 and lies in the direction of $\underset{(n\times 1)}{\mathbf{x}}$ [3, p. 51].

Definition 2.1.13 (Angle). *The* **angle** θ *between two vectors* $\underset{(n\times 1)}{\mathbf{x}}$ *and* $\underset{(n\times 1)}{\mathbf{y}}$ *in a plane, both having* n *entries, is defined from*

$$\cos(\theta) = \frac{\underset{(1\times n)}{\mathbf{x}'} \cdot \underset{(n\times 1)}{\mathbf{y}}}{L_{\mathbf{x}} L_{\mathbf{y}}} = \frac{\underset{(1\times n)}{\mathbf{x}'} \cdot \underset{(n\times 1)}{\mathbf{y}}}{\sqrt{\underset{(1\times n)}{\mathbf{x}'} \cdot \underset{(n\times 1)}{\mathbf{x}}}\sqrt{\underset{(1\times n)}{\mathbf{y}'} \cdot \underset{(n\times 1)}{\mathbf{y}}}}$$

or

$$\cos(\theta) = \frac{\underset{(1\times n)}{\mathbf{y}'} \cdot \underset{(n\times 1)}{\mathbf{x}}}{L_{\mathbf{x}} L_{\mathbf{y}}} = \frac{\underset{(1\times n)}{\mathbf{y}'} \cdot \underset{(n\times 1)}{\mathbf{x}}}{\sqrt{\underset{(1\times n)}{\mathbf{x}'} \cdot \underset{(n\times 1)}{\mathbf{x}}}\sqrt{\underset{(1\times n)}{\mathbf{y}'} \cdot \underset{(n\times 1)}{\mathbf{y}}}}$$

[3, pp. 52-53, 85].

Definition 2.1.14 (Perpendicular). *When the angle between two vectors* $\underset{(n\times 1)}{\mathbf{x}}, \underset{(n\times 1)}{\mathbf{y}}$ *is* $\theta = 90°$ *or* $\theta = 270°$, *we say that* $\underset{(n\times 1)}{\mathbf{x}}$ and $\underset{(n\times 1)}{\mathbf{y}}$ *are* **perpendicular** (orthogonal).

Since $\cos(\theta) = 0$ *only if* $\theta = 90°$ *or* $\theta = 270°$, *the condition becomes*

$$\underset{(n\times 1)}{\mathbf{x}} \text{ and } \underset{(n\times 1)}{\mathbf{y}} \text{ are perpendicular if } \underset{(1\times n)}{\mathbf{x}'} \cdot \underset{(n\times 1)}{\mathbf{y}} = \underset{(1\times n)}{\mathbf{y}'} \cdot \underset{(n\times 1)}{\mathbf{x}} = 0$$

We write $\underset{(n\times 1)}{\mathbf{x}} \perp \underset{(n\times 1)}{\mathbf{y}}$ [3, pp. 53, 86].

Result 2.1.2.

(a) $\underset{(n\times 1)}{\mathbf{z}}$ *is perpendicular to every vector if and only if* $\underset{(n\times 1)}{\mathbf{z}} = \underset{(n\times 1)}{\mathbf{0}}$.

(b) *If* $\underset{(n\times 1)}{\mathbf{z}}$ *is perpendicular to each vector* $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_p$ *then* $\underset{(n\times 1)}{\mathbf{z}}$ *is perpendicular to the* $\operatorname{span}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_p)$.

(c) *Mutually perpendicular vectors are linearly independent.*

[3, p. 86].

2.2 Matrices

Definition 2.2.1 (Matrix). *A* $n \times p$ *dimensional array* $\underset{(n\times p)}{\mathbf{A}}$ *of elements with* n *rows and* p *columns is called a* **matrix**, *and in general, is denoted by a boldfaced, uppercase letter. It is written as*

$$\underset{(n\times p)}{\mathbf{A}} = \underset{(n\times p)}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2k} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jk} & \cdots & a_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} & \cdots & a_{np} \end{bmatrix}} = \underset{(n\times p)}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}}$$

$j = 1,2, \dots, n, k = 1,2, \dots, p$. *Or more compactly as*

$$\underset{(n\times p)}{\mathbf{A}} = \{a_{jk}\}$$

where the index j *refers to the row and the index* k *refers to the column.*

In our work, the matrix elements will be in $\mathbb{R}$ or functions taking on values in $\mathbb{R}$ [3, pp. 54, 87-88].

Definition 2.2.2 (Matrix Transpose). *A $p \times n$ dimensional array $\underset{(p\times n)}{\mathbf{A}'}$ of elements with p rows and n columns is called a* **matrix transpose***,*

$$\underset{(p\times n)}{\mathbf{A}'} = \underset{(p\times n)}{\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{j1} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{j2} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{jk} & \cdots & a_{nk} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{jp} & \cdots & a_{np} \end{bmatrix}} = \underset{(p\times n)}{\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{np} \end{bmatrix}}$$

for $j = 1,2,\dots,n$, $k = 1,2,\dots,p$.

The transpose operation $\underset{(p\times n)}{\mathbf{A}'}$ of a matrix changes the columns into rows, so that the first column of $\underset{(n\times p)}{\mathbf{A}}$ becomes the first row of $\underset{(p\times n)}{\mathbf{A}'}$, the second column becomes the second row, and so forth [3, p. 55].

Definition 2.2.3 (Matrix Addition). *Let the matrices $\underset{(n\times p)}{\mathbf{A}}$ and $\underset{(n\times p)}{\mathbf{B}}$ both be of dimension $n \times p$ with arbitrary elements a_{jk} and b_{jk}, $j = 1,2,\dots,n$, $k = 1,2,\dots,p$, respectively. The sum of the matrices $\underset{(n\times p)}{\mathbf{A}}$ and $\underset{(n\times p)}{\mathbf{B}}$ is an $n \times p$ matrix $\underset{(n\times p)}{\mathbf{C}}$, written $\underset{(n\times p)}{\mathbf{C}} = \underset{(n\times p)}{\mathbf{A}} + \underset{(n\times p)}{\mathbf{B}}$, such that an arbitrary element of $\underset{(n\times p)}{\mathbf{C}}$ is given by*

$$c_{jk} = a_{jk} + b_{jk} \qquad j = 1,2,\dots,n \qquad k = 1,2,\dots,p$$

$$\underset{(n\times p)}{\mathbf{C}} = \underset{(n\times p)}{\mathbf{A}} + \underset{(n\times p)}{\mathbf{B}} = \underset{(n\times p)}{\begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1p}+b_{1p} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2p}+b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & \cdots & a_{np}+b_{np} \end{bmatrix}} = \underset{(n\times p)}{\mathbf{A}+\mathbf{B}}$$

Note that the addition of matrices is defined only for matrices of the same dimension [3, p. 88].

Definition 2.2.4 (Scalar Multiplication).

Let c *be an arbitrary scalar and* $\underset{(n\times p)}{\mathbf{A}} = \{a_{jk}\}$. *Then* $\underset{(n\times p)}{c\mathbf{A}} = \underset{(n\times p)}{\mathbf{A}c} = \underset{(n\times p)}{\mathbf{B}} = \{b_{jk}\}$,

where $b_{jk} = ca_{jk} = a_{jk}c$, $j = 1,2,\dots,n$, $k = 1,2,\dots,p$. *That is,*

$$\underset{(n\times p)}{c\mathbf{A}} = \underset{(n\times p)}{\mathbf{A}c} = \underset{(n\times p)}{\begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1p} \\ ca_{21} & ca_{22} & \cdots & ca_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{np} \end{bmatrix}} = \underset{(n\times p)}{\begin{bmatrix} a_{11}c & a_{12}c & \cdots & a_{1p}c \\ a_{21}c & a_{22}c & \cdots & a_{2p}c \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}c & a_{n2}c & \cdots & a_{np}c \end{bmatrix}} = \underset{(n\times p)}{\mathbf{B}}$$

Multiplication of a matrix by a scalar produces a new matrix whose elements are the elements of the original matrix, each multiplied by the scalar [3, pp. 55, 89].

Definition 2.2.5 (Matrix Multiplication). *The product* $\underset{(n\times m)}{\mathbf{A}} \cdot \underset{(m\times p)}{\mathbf{B}}$ *of an* $n \times m$ *matrix* $\underset{(n\times m)}{\mathbf{A}} = \{a_{jk}\}$ *and an* $m \times p$ *matrix* $\underset{(m\times p)}{\mathbf{B}} = \{b_{jk}\}$ *is the* $n \times p$ *matrix* $\underset{(n\times p)}{\mathbf{C}} = \underset{(n\times m)}{\mathbf{A}} \cdot \underset{(m\times p)}{\mathbf{B}} = \{c_{jk}\}$ *whose elements in the* jth *row and* kth *column is the inner product of the* jth *row of* $\underset{(n\times m)}{\mathbf{A}}$ *and the* kth *column of* $\underset{(m\times p)}{\mathbf{B}}$ *or*

$$c_{jk} = (j,k) \text{ element of } \underset{(n\times m)}{\mathbf{A}} \cdot \underset{(m\times p)}{\mathbf{B}} = a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jm}b_{mk} = \sum_{l=1}^{m} a_{jl}b_{lk}$$

for $j = 1,2,\dots,n$, $k = 1,2,\dots,p$ [3, pp. 55-56, 90].

More generally, the matrix product is given by

$$\underset{(n\times m)}{\mathbf{A}} \cdot \underset{(m\times p)}{\mathbf{B}}$$

$$= \underset{(n\times m)}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jm} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}} \cdot \underset{(m\times p)}{\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2k} & \cdots & b_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} & \cdots & b_{mp} \end{bmatrix}}$$

$$= \underset{(n\times p)}{\begin{array}{lc} & \begin{matrix} \text{Column } 1 & \text{Column } 2 & \cdots & \text{Column } k & \cdots & \text{Column } p \end{matrix} \\ \begin{matrix} \text{row } 1 \\ \text{row } 2 \\ \vdots \\ \text{row } j \\ \vdots \\ \text{row } n \end{matrix} & \begin{bmatrix} \sum_{l=1}^{m} a_{1l}b_{l1} & \sum_{l=1}^{m} a_{1l}b_{l2} & \cdots & \sum_{l=1}^{m} a_{1l}b_{lk} & \cdots & \sum_{l=1}^{m} a_{1l}b_{lp} \\ \sum_{l=1}^{m} a_{2l}b_{l1} & \sum_{l=1}^{m} a_{2l}b_{l2} & \cdots & \sum_{l=1}^{m} a_{2l}b_{lk} & \cdots & \sum_{l=1}^{m} a_{2l}b_{lp} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sum_{l=1}^{m} a_{jl}b_{l1} & \sum_{l=1}^{m} a_{jl}b_{l2} & \cdots & \sum_{l=1}^{m} a_{jl}b_{lk} & \cdots & \sum_{l=1}^{m} a_{jl}b_{lp} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sum_{l=1}^{m} a_{nl}b_{l1} & \sum_{l=1}^{m} a_{nl}b_{l2} & \cdots & \sum_{l=1}^{m} a_{nl}b_{lk} & \cdots & \sum_{l=1}^{m} a_{nl}b_{lp} \end{bmatrix} \end{array}}$$

$$= \underset{(n\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2k} & \cdots & c_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{j1} & c_{j2} & \cdots & c_{jk} & \cdots & c_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nk} & \cdots & c_{np} \end{bmatrix}} = \underset{(n\times p)}{\mathbf{C}}$$

Note that for the product $\underset{(n\times m)}{\mathbf{A}} \cdot \underset{(m\times p)}{\mathbf{B}}$ to be defined, the column dimension of $\underset{(n\times m)}{\mathbf{A}}$ must equal the row dimension of $\underset{(m\times p)}{\mathbf{B}}$. If that is so, then the row dimension of $\underset{(n\times m)}{\mathbf{A}} \cdot \underset{(m\times p)}{\mathbf{B}}$ equals the row dimension of $\underset{(n\times m)}{\mathbf{A}}$, and the column dimension of $\underset{(n\times m)}{\mathbf{A}} \cdot \underset{(m\times p)}{\mathbf{B}}$ equals the column dimension of $\underset{(m\times p)}{\mathbf{B}}$ [3, pp. 55-56, 90].

Result 2.2.1. *For all matrices* $\mathbf{A}$*,* $\mathbf{B}$*, and* $\mathbf{C}$ *(of equal dimension) and scalars* c *and* d*, the following holds:*

(**a**) $\mathbf{A} - \mathbf{B} = \mathbf{A} + (-1)\mathbf{B}$

(**b**) $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$

(**c**) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

(**d**) $c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$

(**e**) $(c + d)\mathbf{A} = c\mathbf{A} + d\mathbf{A}$

(**f**) $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$

(**g**) $(cd)\mathbf{A} = c(d\mathbf{A})$

(**h**) $(c\mathbf{A})' = c\mathbf{A}'$ (Note $c' = c$)

[3, p. 89].

Result 2.2.2. *For all matrices* $\mathbf{A}$*,* $\mathbf{B}$*, and* $\boldsymbol{C}$ *(of dimensions such that the indicated products are defined) and a scalar* c,

(**a**) $c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B}$

(**b**) $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$

(**c**) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

(**d**) $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$

(**e**) $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

[3, p. 91].

Definition 2.2.6 (Zero Matrix). $\underset{(n\times p)}{\mathbf{0}}$ **matrix** *is a rectangular array of* $0's$, *of arbitrary dimension* $n \times p$. *It is written as*

$$\underset{(n\times p)}{\mathbf{0}} = \underset{(n\times p)}{\begin{bmatrix} 0_{11} & 0_{12} & \cdots & 0_{1p} \\ 0_{21} & 0_{22} & \cdots & 0_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n1} & 0_{n2} & \cdots & 0_{np} \end{bmatrix}} = \underset{(n\times p)}{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}}$$

Note that the notation for the $\underset{(p\times 1)}{\mathbf{0}}$ **vector** is similar; but the dimension makes the context clear.

Definition 2.2.7 (Square Matrix). *If an arbitrary matrix* $\underset{(p\times p)}{\mathbf{A}}$ *has the same number of rows and columns, say dimension* $p \times p$, *then* $\underset{(p\times p)}{\mathbf{A}}$ *is called a* **square matrix**. *It is written as*

$$\underset{(p\times p)}{\mathbf{A}} = \underset{(p\times p)}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}} = \{a_{ik}\}$$

for $i = 1,2,\dots,p$ *rows and* $k = 1,2,\dots,p$ *columns* [3, p. 90].

Definition 2.2.8 (Symmetrix Matrix). *Let* $\underset{(p\times p)}{\mathbf{A}} = \{a_{ik}\}$ *be a* $p \times p$ *(square) matrix. Then* $\underset{(p\times p)}{\mathbf{A}}$ *is said to be* a **symmetric matrix** *if* $\underset{(p\times p)}{\mathbf{A}} = \underset{(p\times p)}{\mathbf{A}'}$. *That is,* $\underset{(p\times p)}{\mathbf{A}}$ *is symmetric if* $a_{ik} = a_{ki}\ \forall\ i = 1,2,\dots,p, k = 1,2,\dots,p$ [3, p. 90].

Definition 2.2.9 (Determinant). *The* **determinant** *of a square* $p \times p$ *matrix* $\underset{(p\times p)}{\mathbf{A}}$*, denoted by* $|\mathbf{A}|$*, is the scalar*

$$|\mathbf{A}| = a_{11} \qquad\qquad \text{if } p = 1$$

$$|\mathbf{A}| = \sum_{k=1}^{p} a_{1k}|\mathbf{A}_{1k}|(-1)^{1+k} \qquad \text{if } p > 1$$

where $\mathbf{A}_{1k}$ *is the* $(p-1)\times(p-1)$ *matrix obtained by deleting the first row and* k*th column of* $\underset{(p\times p)}{\mathbf{A}}$*. Also,*

$$|\mathbf{A}| = \sum_{k=1}^{p} a_{ik}|\mathbf{A}_{ik}|(-1)^{i+k} \qquad \text{if } p > 1$$

with the i*th row in place of the first row* [3, p. 93].

Definition 2.2.10 (Identity Matrix). *The* $p \times p$ **identity matrix***, denoted by* $\underset{(p\times p)}{\mathbf{I}}$*, is the square matrix with ones on the main* $(NW - SE)$ *diagonal and zeros elsewhere. It is written as*

$$\underset{(p\times p)}{\mathbf{I}} = \underset{(p\times p)}{\begin{bmatrix} 1_{11} & 0 & \cdots & 0 \\ 0 & 1_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1_{pp} \end{bmatrix}} = \underset{(p\times p)}{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}$$

[3, p. 90].

The matrix $\underset{(p\times p)}{\mathbf{I}}$ acts like 1 in ordinary multiplication $(1 \cdot a = a \cdot 1 = a)$

$$\underset{(p\times p)}{\mathbf{I}} \cdot \underset{(p\times p)}{\mathbf{A}} = \underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times p)}{\mathbf{I}} = \underset{(p\times p)}{\mathbf{A}} \text{ for any } \underset{(p\times p)}{\mathbf{A}}$$

so it is called the identity matrix [3, p. 58].

Remark 2.2.1. There are several important differences between the algebra of matrices and the algebra of real numbers. Two of these differences are as follows:

1. Matrix multiplication is, in general, not commutative. That is, in general,

$$\underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times p)}{\mathbf{B}} \neq \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times p)}{\mathbf{A}}$$

2. Let $\underset{(n\times p)}{\mathbf{0}}$ denote the zero matrix, that is, the matrix with zero for every element. In the algebra of real numbers, if the product of two numbers, ab, is zero, then $a = 0$ or $b = 0$. In matrix algebra, however, the product of two nonzero matrices may be the zero matrix. Hence,

$$\underset{(n\times m)}{\mathbf{A}} \cdot \underset{(m\times p)}{\mathbf{B}} = \underset{(n\times p)}{\mathbf{0}}$$

   does **not** imply that $\underset{(n\times m)}{\mathbf{A}} = \underset{(n\times m)}{\mathbf{0}}$ or $\underset{(m\times p)}{\mathbf{B}} = \underset{(m\times p)}{\mathbf{0}}$. It is true, however, that if either

$$\underset{(n\times m)}{\mathbf{A}} = \underset{(n\times m)}{\mathbf{0}} \text{ or } \underset{(m\times p)}{\mathbf{B}} = \underset{(m\times p)}{\mathbf{0}}, \text{ then } \underset{(n\times m)}{\mathbf{A}} \cdot \underset{(m\times p)}{\mathbf{B}} = \underset{(n\times p)}{\mathbf{0}}$$

   [3, pp. 58, 92].

Definition 2.2.11 (Row Rank and Column Rank)**.** *The* **row rank** *of a matrix is the maximum number of linearly independent rows, considered as vectors. The* **column rank** *of a matrix is the rank of its set of columns, considered as vectors* [3, p. 94].

Result 2.2.3 (Rank of a Matrix)**.** *The row rank and the column rank of a matrix are equal. Thus, the* **rank of a matrix** *is either the row rank or the column rank* [3, p. 94].

Definition 2.2.12 (Nonsingular). *A square matrix* $\underset{(p\times p)}{\mathbf{A}}$ *is* **nonsingular** *if*

$$\underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times 1)}{\mathbf{0}}$$

implies

$$\underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times 1)}{\mathbf{0}}.$$

If a matrix fails to be nonsingular, it is called **singular**. *Equivalently, a square matrix is nonsingular if its rank is equal to the number of rows (or columns) it has.*

Note that $\underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times 1)}{\mathbf{x}} = x_1 \underset{(p\times 1)}{\mathbf{a}_1} + x_2 \underset{(p\times 1)}{\mathbf{a}_2} + \cdots + x_k \underset{(p\times 1)}{\mathbf{a}_k} + \cdots + x_p \underset{(p\times 1)}{\mathbf{a}_p}$, where $x_k \underset{(p\times 1)}{\mathbf{a}_k}$ is the kth column of $\underset{(p\times p)}{\mathbf{A}}$, so that the condition of nonsingularity is just the statement that the columns of $\underset{(p\times p)}{\mathbf{A}}$ are linearly independent [3, p. 95].

Definition 2.2.13 (Inverse). *Let* $\underset{(p\times p)}{\mathbf{A}}$ *be a nonsingular square matrix of dimension* $p \times p$. *Then there is a unique* $p \times p$ *matrix* $\underset{(p\times p)}{\mathbf{B}}$ *such that*

$$\underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times p)}{\mathbf{B}} = \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times p)}{\mathbf{A}} = \underset{(p\times p)}{\mathbf{I}}$$

where $\underset{(p\times p)}{\mathbf{I}}$ *is the* $p \times p$ *identity matrix. Then* $\underset{(p\times p)}{\mathbf{B}}$ *is called the* **inverse** *of* $\underset{(p\times p)}{\mathbf{A}}$ *and is denoted by* $\underset{(p\times p)}{\mathbf{A}^{-1}}$ [3, p. 95].

Result 2.2.4. *For a square matrix* $\underset{(p\times p)}{\mathbf{A}}$ *of dimension* $p \times p$*, the following are equivalent:*

(a) $\underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times 1)}{\mathbf{0}}$ *implies* $\underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times 1)}{\mathbf{0}} \left(\underset{(p\times p)}{\mathbf{A}} \textit{ is nonsingular}\right)$.

(b) $|\mathbf{A}| \neq 0$ *where* $(|\cdot| \textit{ denotes the determinant operator})$.

(c) *There exists a matrix* $\underset{(p\times p)}{\mathbf{A}^{-1}}$ *such that* $\underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times p)}{\mathbf{A}^{-1}} = \underset{(p\times p)}{\mathbf{A}^{-1}} \cdot \underset{(p\times p)}{\mathbf{A}} = \underset{(p\times p)}{\mathbf{I}}$.

[3, p. 96].

Result 2.2.5. *Let* $\underset{(p\times p)}{\mathbf{A}}$ *and* $\underset{(p\times p)}{\mathbf{B}}$ *be* $p \times p$ *square matrices, and let the indicated inverses exist. Then the following hold:*

(a) $\underset{(p\times p)}{(\mathbf{A}^{-1})'} = \underset{(p\times p)}{(\mathbf{A}')^{-1}}$

(b) $\underset{(p\times p)}{(\mathbf{AB})^{-1}} = \underset{(p\times p)}{\mathbf{B}^{-1}} \cdot \underset{(p\times p)}{\mathbf{A}^{-1}}$

[3, p. 96].

Definition 2.2.14 (Trace). *Let* $\underset{(p\times p)}{\mathbf{A}} = \{a_{ik}\}$ *be a* $p \times p$ *square matrix. The* **trace** *of the matrix* $\underset{(p\times p)}{\mathbf{A}}$*, written* $\mathrm{tr}(\mathbf{A})$ *is the sum of the diagonal elements; that is,*

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{p} a_{ii}$$

[3, p. 96].

Result 2.2.6. *Let* $\underset{(p\times p)}{\mathbf{A}}$ *and* $\underset{(p\times p)}{\mathbf{B}}$ *be* $p \times p$ *square matrices,* $\underset{(p\times p)}{\mathbf{B}^{-1}}$ *exist, and* c *be a scalar.*

($\mathbf{a}$) $\mathrm{tr}(c\mathbf{A}) = c\,\mathrm{tr}(\mathbf{A})$

($\mathbf{b}$) $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$

($\mathbf{c}$) $\mathrm{tr}(\mathbf{B}^{-1}\mathbf{AB}) = \mathrm{tr}(\mathbf{A})$

[3, p. 97].

Definition 2.2.15 (Orthogonal)**.** *A square matrix* $\underset{(p\times p)}{\mathbf{A}}$ *is said to be* **orthogonal** *if its rows*

$$\underset{(p\times 1)}{\mathbf{a}_r} = \begin{bmatrix} a_{r1} \\ a_{r2} \\ \vdots \\ a_{rp} \end{bmatrix}$$

for $r = 1,2,\dots,p$, *considered as vectors, are mutually perpendicular,*

$$\underset{(1\times p)}{\mathbf{a}_r'} \cdot \underset{(p\times 1)}{\mathbf{a}_s} = 0 \; for \; r \neq s$$

and have unit lengths

$$\underset{(1\times p)}{\mathbf{a}_r'} \cdot \underset{(p\times 1)}{\mathbf{a}_r} = 1$$

that is,

$$\underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times p)}{\mathbf{A}'} = \underset{(p\times p)}{\mathbf{I}},$$

and its columns

$$\underset{(p\times 1)}{\mathbf{a}_i} = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{pi} \end{bmatrix}$$

for $i = 1,2,\dots,p$, *considered as vectors, are mutually perpendicular,*

$$\underset{(1\times p)}{\mathbf{a}_i'} \cdot \underset{(p\times 1)}{\mathbf{a}_k} = 0 \; for \; i \neq k$$

and have unit lengths

$$\underset{(1\times p)}{\mathbf{a}_i'} \cdot \underset{(p\times 1)}{\mathbf{a}_i} = 1$$

that is,

$$\underset{(p\times p)}{\mathbf{A}'} \cdot \underset{(p\times p)}{\mathbf{A}} = \underset{(p\times p)}{\mathbf{I}}$$

[3, pp. 59, 97].

Result 2.2.7. *A square matrix* $\underset{(p\times p)}{\mathbf{A}}$ *is orthogonal if and only if* $\underset{(p\times p)}{\mathbf{A}^{-1}} = \underset{(p\times p)}{\mathbf{A}'}$. *For an orthogonal matrix,* $\underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times p)}{\mathbf{A}'} = \underset{(p\times p)}{\mathbf{A}'} \cdot \underset{(p\times p)}{\mathbf{A}} = \underset{(p\times p)}{\mathbf{I}}$, *so, the rows and columns are also mutually perpendicular* [3, pp. 59, 97].

Definition 2.2.16 (Eigenvalues). *Let* $\underset{(p\times p)}{\mathbf{A}}$ *be a* $p\times p$ *square matrix and* $\underset{(p\times p)}{\mathbf{I}}$ *be the* $p\times p$ *identity matrix. Then the scalars* $\lambda_1,\lambda_2,\dots,\lambda_p$ *satisfying the polynomial equation* $|\mathbf{A}-\lambda\mathbf{I}|=0$ *are called the* **eigenvalues** *(or characteristic roots) of a matrix* $\underset{(p\times p)}{\mathbf{A}}$. *The equation* $|\mathbf{A}-\lambda\mathbf{I}|=0$ *(as a function of* λ*) is called the* **characteristic equation** [3, p. 97].

Definition 2.2.17 (Eigenvector). *Let* $\underset{(p\times p)}{\mathbf{A}}$ *be a square matrix of dimension* $p\times p$ *and let* λ *be an eigenvalue of* $\underset{(p\times p)}{\mathbf{A}}$. *If* $\underset{(p\times 1)}{\mathbf{x}}$ *is a nonzero vector* $\left(\underset{(p\times 1)}{\mathbf{x}}\neq\underset{(p\times 1)}{\mathbf{0}}\right)$ *such that*

$$\underset{(p\times p)}{\mathbf{A}}\cdot\underset{(p\times 1)}{\mathbf{x}}=\lambda\cdot\underset{(p\times 1)}{\mathbf{x}}$$

then $\underset{(p\times 1)}{\mathbf{x}}$ *is said to be an* **eigenvector** *(characteristic vector) of the matrix* $\underset{(p\times p)}{\mathbf{A}}$ *associated with the eigenvalue* λ [3, p. 98].

An equivalent condition for λ to be a solution of the eigenvalue-eigenvector equation is $|\mathbf{A}-\lambda\mathbf{I}|=0$. This follows because the statement that $\underset{(p\times p)}{\mathbf{A}}\cdot\underset{(p\times 1)}{\mathbf{x}}=\underset{(p\times 1)}{\lambda\mathbf{x}}$ for some λ and $\underset{(p\times 1)}{\mathbf{x}}\neq\underset{(p\times 1)}{\mathbf{0}}$ implies that

$$\underset{(p\times 1)}{\mathbf{0}}=\underset{(p\times p)}{(\mathbf{A}-\lambda\mathbf{I})}\cdot\underset{(p\times 1)}{\mathbf{x}}=x_1\cdot\underset{(p\times 1)}{\mathrm{col}_1(\mathbf{A}-\lambda\mathbf{I})}+\cdots+x_p\cdot\underset{(p\times 1)}{\mathrm{col}_p(\mathbf{A}-\lambda\mathbf{I})}$$

That is, the columns of $\underset{(p\times p)}{(\mathbf{A}-\lambda\mathbf{I})}$ are linearly dependent so, by Result 2.2.4. (b), $|\mathbf{A}-\lambda\mathbf{I}|=0$, as asserted [3, p. 98].

Ordinarily, we normalize $\underset{(p\times 1)}{\mathbf{x}}$ so that it has length unity. It is convenient to denote normalized eigenvectors by

$$\underset{(p\times 1)}{\mathbf{e}} = L_{\mathbf{x}}^{-1} \cdot \underset{(p\times 1)}{\mathbf{x}} = \frac{\underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times 1)}{L_{\mathbf{x}}}} = \frac{\underset{(p\times 1)}{\mathbf{x}}}{\sqrt{\underset{(1\times n)}{\mathbf{x}'} \cdot \underset{(n\times 1)}{\mathbf{x}}}}$$

and we do so in what follows [3, pp. 60, 99].

Definition 2.2.18 (Eigenvalue-Eigenvector Pairs). *Let* $\underset{(p\times p)}{\mathbf{A}}$ *be a* $p \times p$ *square symmetric matrix. Then* $\underset{(p\times p)}{\mathbf{A}}$ *has* p **eigenvalue-eigenvector pairs**-*namely,*

$$\left(\lambda_1, \underset{(p\times 1)}{\mathbf{e}_1}\right), \left(\lambda_2, \underset{(p\times 1)}{\mathbf{e}_2}\right), \dots, \left(\lambda_i, \underset{(p\times 1)}{\mathbf{e}_i}\right), \dots, \left(\lambda_p, \underset{(p\times 1)}{\mathbf{e}_p}\right).$$

Let the normalized eigenvectors be the columns of another matrix

$$\underset{(p\times p)}{\mathbf{E}} = \underset{(p\times p)}{\begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix}}$$

where the columns of the $\underset{(p\times p)}{\mathbf{E}}$ are mutually perpendicular

$$\underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times 1)}{\mathbf{e}_k} = 0 \text{ for } i \neq k$$

and have unit lengths

$$\underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times 1)}{\mathbf{e}_i} = 1$$

that is,

$$\underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times p)}{\mathbf{E}} = \underset{(p\times p)}{\mathbf{I}}$$

And the rows of $\underset{(p\times p)}{\mathbf{E}}$ are mutually perpendicular

$$\underset{(1\times p)}{\mathbf{e}'_r} \cdot \underset{(p\times 1)}{\mathbf{e}_s} = 0 \text{ for } r \neq s$$

and have unit lengths

$$\underset{(1\times p)}{\mathbf{e}'_r} \cdot \underset{(p\times 1)}{\mathbf{e}_r} = 1$$

that is,

$$\underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{E}'} = \underset{(p\times p)}{\mathbf{I}}$$

Thus, $\underset{(p\times p)}{\mathbf{E}}$ is orthogonal making

$$\underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{E}'} = \underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times p)}{\mathbf{E}} = \underset{(p\times p)}{\mathbf{I}}$$

and

$$\underset{(p\times p)}{\mathbf{E}^{-1}} = \underset{(p\times p)}{\mathbf{E}'}$$

Let us demonstrate,

$$\underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{E}'}$$

$$= \underset{(p\times p)}{\begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix}} \cdot \underset{(p\times p)}{\begin{bmatrix} e_{11} & e_{21} & \cdots & e_{p1} \\ e_{12} & e_{22} & \cdots & e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{2p} & \cdots & e_{pp} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \mathbf{e}'_1\mathbf{e}_1 = 1 & \mathbf{e}'_1\mathbf{e}_2 = 0 & \cdots & \mathbf{e}'_1\mathbf{e}_p = 0 \\ \mathbf{e}'_2\mathbf{e}_1 = 0 & \mathbf{e}'_2\mathbf{e}_2 = 1 & \cdots & \mathbf{e}'_2\mathbf{e}_p = 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{e}'_p\mathbf{e}_1 = 0 & \mathbf{e}'_p\mathbf{e}_2 = 0 & \cdots & \mathbf{e}'_p\mathbf{e}_p = 1 \end{bmatrix}}, \{\text{rows perpendicular}\}$$

$$= \underset{(p\times p)}{\begin{bmatrix} 1_1 & 0 & \cdots & 0 \\ 0 & 1_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1_p \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}$$

$$= \underset{(p\times p)}{\mathbf{I}}$$

and

$$\underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times p)}{\mathbf{E}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} e_{11} & e_{21} & \cdots & e_{p1} \\ e_{12} & e_{22} & \cdots & e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{2p} & \cdots & e_{pp} \end{bmatrix}} \underset{(p\times p)}{\begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \mathbf{e}_1'\mathbf{e}_1 = 1 & \mathbf{e}_1'\mathbf{e}_2 = 0 & \cdots & \mathbf{e}_1'\mathbf{e}_p = 0 \\ \mathbf{e}_2'\mathbf{e}_1 = 0 & \mathbf{e}_2'\mathbf{e}_2 = 1 & \cdots & \mathbf{e}_2'\mathbf{e}_p = 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{e}_p'\mathbf{e}_1 = 0 & \mathbf{e}_p'\mathbf{e}_2 = 0 & \cdots & \mathbf{e}_p'\mathbf{e}_p = 1 \end{bmatrix}}, \{\text{columns perpendicular}\}$$

$$= \underset{(p\times p)}{\begin{bmatrix} 1_1 & 0 & \cdots & 0 \\ 0 & 1_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1_p \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}$$

$$= \underset{(p\times p)}{\mathbf{I}} \blacksquare$$

Note that the eigenvectors are unique unless two or more eigenvalues are equal. Clearly, $\underset{(p\times1)}{\mathbf{e}_1}$, $\underset{(p\times1)}{\mathbf{e}_2}$, …, $\underset{(p\times1)}{\mathbf{e}_i}$, …, $\underset{(p\times1)}{\mathbf{e}_p}$ are the (normalized) solutions of the equations $\underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times1)}{\mathbf{e}_i} = \lambda_i \cdot \underset{(p\times1)}{\mathbf{e}_i}$ for $i = 1,2,\dots,p$ [3, pp. 60-61, 65].

Definition 2.2.19 (Quadratic Form)**.** *A* **quadratic form** $\underset{(1\times1)}{Q(\mathbf{x})}$ *in the* p *variables* $x_1, x_2, \dots, x_p$ *is* $\underset{(1\times1)}{Q(\mathbf{x})} = \underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times1)}{\mathbf{x}}$, *where* $\underset{(1\times p)}{\mathbf{x}'} = \underset{(1\times p)}{[x_1, x_2, \dots, x_p]}$ *and* $\underset{(p\times p)}{\mathbf{A}}$ *is a* $p \times p$ *symmetric matrix.*

Note that a quadratic form can be written as

$$\underset{(1\times1)}{Q(\mathbf{x})} = \sum_{i=1}^{p} \sum_{k=1}^{p} a_{ik} x_i x_k$$

Because $\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times1)}{\mathbf{x}}$ has only squared terms x_i^2 and product terms $x_i x_k$, it is called a quadratic form [3, pp. 62, 99].

If $\exists \underset{(p\times1)}{\mathbf{x}} \neq \underset{(p\times1)}{\mathbf{0}}$ and a $p \times p$ symmetric matrix $\underset{(p\times p)}{\mathbf{A}}$ where

$$\underset{(1\times1)}{0} = \underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times1)}{\mathbf{x}}$$

then the matrix $\underset{(p\times p)}{\mathbf{A}}$ and the quadratic form are said to be **positive semi-definite.** If

$$\underset{(1\times1)}{0} < \underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times1)}{\mathbf{x}}$$

$\forall \underset{(p\times1)}{\mathbf{x}} \neq \underset{(p\times1)}{\mathbf{0}}$ then the $p \times p$ symmetric matrix $\underset{(p\times p)}{\mathbf{A}}$ and the quadratic form are said to be **positive definite** [3, p. 62].

In addition, when $\underset{(p\times p)}{\mathbf{A}}$ is positive definite the quadratic form can be interpreted as a squared distance [3, p. 64]. If the quadratic form and the matrix $\underset{(p\times p)}{\mathbf{A}}$ are positive semi-definite or positive definite they are said to be nonnegative definite [3, p. 62].

Results involving quadratic forms and symmetric matrices are, in many cases, a direct consequence of an expansion for symmetric matrices known as the **spectral decomposition.** That is, any symmetric square matrix can be reconstructed from its eigenvalues and eigenvectors. The particular expression reveals the relative importance of each pair according to the relative size of the eigenvalue and the direction of the eigenvector [3, pp. 61, 99].

Result 2.2.8 (Spectral Decomposition)**.** *The* **Spectral Decomposition**. *Let* $\underset{(p\times p)}{\mathbf{A}}$ *be a* $p \times p$ *symmetric matrix. Then* $\underset{(p\times p)}{\mathbf{A}}$ *can be expressed in terms of its* p *eigenvalue-eigenvector pairs* $\left(\lambda_i, \underset{(p\times 1)}{\mathbf{e}_i}\right)$ *as*

$$\underset{(p\times p)}{\mathbf{A}} = \sum_{i=1}^{p} \lambda_i \cdot \underset{(p\times 1)}{\mathbf{e}_i} \cdot \underset{(1\times p)}{\mathbf{e}_i'} = \lambda_1 \cdot \underset{(p\times 1)}{\mathbf{e}_1} \cdot \underset{(1\times p)}{\mathbf{e}_1'} + \cdots + \lambda_p \cdot \underset{(p\times 1)}{\mathbf{e}_p} \cdot \underset{(1\times p)}{\mathbf{e}_p'}$$

where $\lambda_1, \lambda_1, \ldots, \lambda_p$ *are the eigenvalues of* $\underset{(p\times p)}{\mathbf{A}}$ *and* $\underset{(p\times 1)}{\mathbf{e}_1}, \underset{(p\times 1)}{\mathbf{e}_2}, \ldots, \underset{(p\times 1)}{\mathbf{e}_p}$ *are the associated normalized eigenvectors* [3, pp. 61, 100].

Using the spectral decomposition, we can easily show that a $p \times p$ symmetric matrix $\underset{(p\times p)}{\mathbf{A}}$ is a positive definite matrix if and only if every eigenvalue of $\underset{(p\times p)}{\mathbf{A}}$ is positive [$\lambda_i > 0 \; \forall \; i$]. Similarly, $\underset{(p\times p)}{\mathbf{A}}$ is a positive semi-definite matrix if and only if $\exists \; \lambda_i = 0$ and the other eigenvalues are positive [5, pp. 212, 549].

The spectral decomposition allows us to express the inverse of a square matrix in terms of its eigenvalues and eigenvectors, and this leads to a useful **square-root matrix.**

Let $\underset{(p\times p)}{\mathbf{A}}$ be a $p \times p$ positive definite matrix with the spectral decomposition

$$\underset{(p\times p)}{\mathbf{A}} = \sum_{i=1}^{p} \lambda_i \cdot \underset{(p\times 1)}{\mathbf{e}_i} \cdot \underset{(1\times p)}{\mathbf{e}_i'} .$$

Let $\underset{(p\times p)}{\mathbf{E}}$ be a $p \times p$ orthogonal matrix with columns equal to the normalized eigenvectors of $\underset{(p\times p)}{\mathbf{A}}$,

$$\underset{(p\times p)}{\mathbf{E}} = \underset{(p\times p)}{\begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix}}$$

let and $\underset{(p\times p)}{\mathbf{\Lambda}}$ be the the diagonal matrix of eigenvalues of $\underset{(p\times p)}{\mathbf{A}}$,

$$\underset{(p\times p)}{\mathbf{\Lambda}} = \underset{(p\times p)}{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}}$$

with inverse,

$$\underset{(p\times p)}{\mathbf{\Lambda}^{-1}} = \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_p} \end{bmatrix}}$$

Then

$$\underset{(p\times p)}{\mathbf{A}} = \sum_{i=1}^{p} \lambda_i \cdot \underset{(p\times 1)}{\mathbf{e}_i} \cdot \underset{(1\times p)}{\mathbf{e}_i'} = \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times p)}{\mathbf{E}'}$$

With inverse

$$\underset{(p\times p)}{\mathbf{A}^{-1}} = \sum_{i=1}^{p} \frac{1}{\lambda_i} \cdot \underset{(p\times 1)}{\mathbf{e}_i} \cdot \underset{(1\times p)}{\mathbf{e}_i'} = \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{-1}} \cdot \underset{(p\times p)}{\mathbf{E}'}$$

since,

$$\underset{(p\times p)}{\mathbf{A}^{-1}} \cdot \underset{(p\times p)}{\mathbf{A}}$$

$$= \left[\underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{-1}} \cdot \underset{(p\times p)}{\mathbf{E}'}\right]\left[\underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times p)}{\mathbf{E}'}\right]$$

$$= \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{-1}} \cdot \left[\underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times p)}{\mathbf{E}}\right] \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times p)}{\mathbf{E}'}, \{\textbf{Result 2.2.2 (b)}\}$$

$$= \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{-1}} \cdot \underset{(p\times p)}{\mathbf{I}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times p)}{\mathbf{E}'}, \{\mathbf{E} \text{ is orthogonal}, \mathbf{E}^{-1} = \mathbf{E}'\}$$

$$= \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{-1}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times p)}{\mathbf{E}'}$$

$$= \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{I}} \cdot \underset{(p\times p)}{\mathbf{E}'}, \{\mathbf{\Lambda}^{-1} \text{ is inverse of } \mathbf{\Lambda}\}$$

$$= \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{E}'}$$

$$= \underset{(p\times p)}{\mathbf{I}}, \{\mathbf{E} \text{ is orthogonal}, \mathbf{E}^{-1} = \mathbf{E}'\}$$

and

$$\underset{(p\times p)}{\mathbf{A}} \cdot \underset{(p\times p)}{\mathbf{A}^{-1}}$$

$$= \left[\underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times p)}{\mathbf{E}'}\right]\left[\underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{-1}} \cdot \underset{(p\times p)}{\mathbf{E}'}\right]$$

$$= \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \left[\underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times p)}{\mathbf{E}}\right] \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{-1}} \cdot \underset{(p\times p)}{\mathbf{E}'}, \{\mathbf{Result\ 2.2.2\ (b)}\}$$

$$= \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times p)}{\mathbf{I}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{-1}} \cdot \underset{(p\times p)}{\mathbf{E}'}, \{\mathbf{E} \text{ is orthogonal}, \mathbf{E}^{-1} = \mathbf{E}'\}$$

$$= \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{-1}} \cdot \underset{(p\times p)}{\mathbf{E}'}$$

$$= \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{I}} \cdot \underset{(p\times p)}{\mathbf{E}'}, \{\mathbf{\Lambda}^{-1} \text{ is inverse of } \mathbf{\Lambda}\}$$

$$= \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{E}'}$$

$$= \underset{(p\times p)}{\mathbf{I}}, \{\mathbf{E} \text{ is orthogonal}, \mathbf{E}^{-1} = \mathbf{E}'\} \blacksquare$$

[3, pp. 65-66].

Definition 2.2.20 (Square-Root Matrix). *Let* $\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}$ *denote the diagonal matrix with* $\sqrt{\lambda_i}$ *as the* ith *diagonal element. Then the* **square-root matrix**, *of a positive definite matrix* $\underset{(p\times p)}{\mathbf{A}}$ *is given by*

$$\underset{(p\times p)}{\mathbf{A}^{1/2}} = \sum_{i=1}^{p} \sqrt{\lambda_i} \cdot \underset{(p\times 1)}{\mathbf{e}_i} \cdot \underset{(1\times p)}{\mathbf{e}_i'} = \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{1/2}} \cdot \underset{(p\times p)}{\mathbf{E}'}$$

[3, p. 66].

Result 2.2.9. *The square-root matrix* $\underset{(p\times p)}{\mathrm{A}^{1/2}}$ *has the following properties:*

(**a**) $\left(\underset{(p\times p)}{\mathbf{A}^{1/2}}\right)' = \underset{(p\times p)}{\mathbf{A}^{1/2}}$, $(that\ is, \mathbf{A}^{1/2}\ is\ symmetric)$

(**b**) $\underset{(p\times p)}{\mathbf{A}^{1/2}} \cdot \underset{(p\times p)}{\mathbf{A}^{1/2}} = \underset{(p\times p)}{\mathbf{A}}, \{\left[\underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{1/2}} \cdot \underset{(p\times p)}{\mathbf{E}'}\right]\left[\underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{1/2}} \cdot \underset{(p\times p)}{\mathbf{E}'}\right]\}$

(**c**) $\left(\underset{(p\times p)}{\mathbf{A}^{1/2}}\right)^{-1} = \sum_{i=1}^{p} \frac{1}{\sqrt{\lambda_i}} \cdot \underset{(p\times 1)}{\mathbf{e}_i} \cdot \underset{(1\times p)}{\mathbf{e}_i'} = \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{-1/2}} \cdot \underset{(p\times p)}{\mathbf{E}'}$, *where* $\underset{(p\times p)}{\mathbf{\Lambda}^{-1/2}}$ *is a diagonal matrix with* $1/\sqrt{\lambda_i}$ *as the* ith *diagonal element.*

(**d**) $\underset{(p\times p)}{\mathbf{A}^{1/2}} \cdot \underset{(p\times p)}{\mathbf{A}^{-1/2}} = \underset{(p\times p)}{\mathbf{A}^{-1/2}} \cdot \underset{(p\times p)}{\mathbf{A}^{1/2}} = \underset{(p\times p)}{\mathbf{I}}$ $(inverse)$, *and* $\underset{(p\times p)}{\mathbf{A}^{-1/2}} \cdot \underset{(p\times p)}{\mathbf{A}^{-1/2}} = \underset{(p\times p)}{\mathbf{A}^{-1}}$, *where* $\underset{(p\times p)}{\mathbf{A}^{-1/2}} = \left(\underset{(p\times p)}{\mathbf{A}^{1/2}}\right)^{-1}$.

[3, p. 66].

Theorem 2.2.1 (Maximization of Quadratic Forms for Points on the Unit Sphere)**.**

Let $\underset{(p\times p)}{\mathbf{B}}$ *be a positive definite matrix with eigenvalues* $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$ *and associated normalized eigenvectors* $\underset{(p\times 1)}{\mathbf{e}_1}, \underset{(p\times 1)}{\mathbf{e}_2}, \ldots, \underset{(p\times 1)}{\mathbf{e}_p}$. *Then*

$$\max_{\underset{(p\times 1)}{\mathbf{x}} \neq \underset{(p\times 1)}{\mathbf{0}}} \frac{\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times 1)}{\mathbf{x}}} = \lambda_1 \qquad \left(attained\ when\ \underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times 1)}{\mathbf{e}_1}\right)$$

$$\min_{\underset{(p\times 1)}{\mathbf{x}} \neq \underset{(p\times 1)}{\mathbf{0}}} \frac{\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times 1)}{\mathbf{x}}} = \lambda_p \qquad \left(attained\ when\ \underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times 1)}{\mathbf{e}_p}\right)$$

Moreover,

$$\max_{\underset{(p\times 1)}{\mathbf{x}} \perp \underset{(p\times 1)}{\mathbf{e}_1}, \underset{(p\times 1)}{\mathbf{e}_2}, \ldots, \underset{(p\times 1)}{\mathbf{e}_k}} \frac{\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times 1)}{\mathbf{x}}} = \lambda_{k+1},$$

$(attained\ when\ \underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times 1)}{\mathbf{e}_{k+1}}, k = 1,2, \ldots, p-1)$

where the symbol $\perp$ *is read "is perpendicular to."*

Proof: Let $\underset{(p\times p)}{\mathbf{E}}$ be the orthogonal matrix whose columns are the eigenvectors

$\underset{(p\times 1)}{\mathbf{e}_1}, \underset{(p\times 1)}{\mathbf{e}_2}, \ldots, \underset{(p\times 1)}{\mathbf{e}_p}$ and $\underset{(p\times p)}{\mathbf{\Lambda}}$ be the diagonal matrix with eigenvalues $\lambda_1, \lambda_1, \ldots, \lambda_p$

along the main diagonal. Let $\underset{(p\times p)}{\mathbf{B}^{1/2}} = \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{1/2}} \cdot \underset{(p\times p)}{\mathbf{E}'}$ (square-root matrix)

and $\underset{(p\times 1)}{\mathbf{y}} = \underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times 1)}{\mathbf{x}}$.

Consequently, $\underset{(p\times 1)}{\mathbf{x}} \neq \underset{(p\times 1)}{\mathbf{0}}$ implies $\underset{(p\times 1)}{\mathbf{y}} \neq \underset{(p\times 1)}{\mathbf{0}}$ because $\underset{(p\times p)}{\mathbf{E}'}$ is an orthogonal matrix and hence has inverse $\underset{(p\times p)}{\mathbf{E}} \left\{ \underset{(p\times p)}{\mathbf{E}'} = \underset{(p\times p)}{\mathbf{E}^{-1}} \right\}$. Thus,

$\underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times 1)}{\mathbf{y}}$. But $\underset{(p\times 1)}{\mathbf{x}}$ is a nonzero vector, and $\underset{(p\times 1)}{\mathbf{0}} \neq \underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times 1)}{\mathbf{y}}$

implies that $\underset{(p\times 1)}{\mathbf{y}} \neq \underset{(p\times 1)}{\mathbf{0}}$.

Thus,

$$\frac{\underset{(1\times p)}{\mathbf{x}'}\cdot\underset{(p\times p)}{\mathbf{B}}\cdot\underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times p)}{\mathbf{x}'}\cdot\underset{(p\times 1)}{\mathbf{x}}}$$

$$=\frac{\underset{(1\times p)}{\mathbf{x}'}\cdot\left[\underset{(p\times p)}{\mathbf{B}^{1/2}}\cdot\underset{(p\times p)}{\mathbf{B}^{1/2}}\right]\cdot\underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times p)}{\mathbf{x}'}\cdot\underset{(p\times 1)}{\mathbf{x}}},\left\{\underset{(p\times p)}{\mathbf{B}^{1/2}}\cdot\underset{(p\times p)}{\mathbf{B}^{1/2}}=\underset{(p\times p)}{\mathbf{B}},\mathbf{Result\ 2.2.9.(b)}\right\}$$

$$=\frac{\underset{(1\times p)}{\mathbf{x}'}\cdot\left[\left[\underset{(p\times p)}{\mathbf{E}}\cdot\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\cdot\underset{(p\times p)}{\mathbf{E}'}\right]\left[\underset{(p\times p)}{\mathbf{E}}\cdot\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\cdot\underset{(p\times p)}{\mathbf{E}'}\right]\right]\cdot\underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times p)}{\mathbf{x}'}\cdot\underset{(p\times 1)}{\mathbf{x}}},$$

$$\left\{\underset{(p\times p)}{\mathbf{B}^{1/2}}=\underset{(p\times p)}{\mathbf{E}}\cdot\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\cdot\underset{(p\times p)}{\mathbf{E}'},\mathbf{Result\ 2.2.9.\ (b)}\right\}$$

$$=\frac{\underset{(1\times p)}{\mathbf{x}'}\cdot\left[\underset{(p\times p)}{\mathbf{E}}\cdot\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\cdot\left[\underset{(p\times p)}{\mathbf{E}'}\cdot\underset{(p\times p)}{\mathbf{E}}\right]\cdot\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\cdot\underset{(p\times p)}{\mathbf{E}'}\right]\cdot\underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times p)}{\mathbf{x}'}\cdot\underset{(p\times 1)}{\mathbf{x}}},\{\mathbf{Result\ 2.2.2\ (b)}\}$$

$$=\frac{\underset{(1\times p)}{\mathbf{x}'}\cdot\left[\underset{(p\times p)}{\mathbf{E}}\cdot\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\cdot\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\cdot\underset{(p\times p)}{\mathbf{E}'}\right]\cdot\underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times p)}{\mathbf{x}'}\cdot\underset{(p\times 1)}{\mathbf{x}}},\left\{\underset{(p\times p)}{\mathbf{E}}\text{ is orthogonal, }\underset{(p\times p)}{\mathbf{E}^{-1}}=\underset{(p\times p)}{\mathbf{E}'}\right\}$$

$$=\frac{\left[\underset{(1\times p)}{\mathbf{x}'}\cdot\underset{(p\times p)}{\mathbf{E}}\right]\left[\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\cdot\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\right]\left[\underset{(p\times p)}{\mathbf{E}'}\cdot\underset{(p\times 1)}{\mathbf{x}}\right]}{\underset{(1\times p)}{\mathbf{x}'}\cdot\underset{(p\times 1)}{\mathbf{x}}},\{\mathbf{Result\ 2.2.2.(b)}\}$$

$$=\frac{\left[\underset{(1\times p)}{\mathbf{x}'}\cdot\underset{(p\times p)}{\mathbf{E}}\right]\left[\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\cdot\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\right]\left[\underset{(p\times p)}{\mathbf{E}'}\cdot\underset{(p\times 1)}{\mathbf{x}}\right]}{\underset{(1\times p)}{\mathbf{x}'}\cdot\underset{(p\times p)}{\mathbf{E}}\cdot\underset{(p\times p)}{\mathbf{E}'}\cdot\underset{(p\times 1)}{\mathbf{x}}},$$

$$\left\{\underset{(p\times p)}{\mathbf{E}}\text{ is orthogonal, }\underset{(p\times p)}{\mathbf{E}^{-1}}=\underset{(p\times p)}{\mathbf{E}'},\underset{(p\times p)}{\mathbf{E}}\cdot\underset{(p\times p)}{\mathbf{E}'}=\underset{(p\times p)}{\mathbf{I}}\right\}$$

$$= \frac{\left[\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{E}}\right]\left[\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\right]\left[\underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times 1)}{\mathbf{x}}\right]}{\left[\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{E}}\right]\left[\underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times 1)}{\mathbf{x}}\right]}, \{\mathbf{Result\ 2.2.2.(b)}\}$$

$$= \frac{\underset{(1\times p)}{\mathbf{y}'}\left[\underset{(p\times p)}{\mathbf{\Lambda}^{1/2}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}^{1/2}}\right]\underset{(p\times 1)}{\mathbf{y}}}{\underset{(1\times p)}{\mathbf{y}'} \cdot \underset{(p\times 1)}{\mathbf{y}}}, \left\{\underset{(p\times 1)}{\mathbf{y}} = \underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times 1)}{\mathbf{x}}, \mathbf{Result\ 2.2.2.(e)}\right\}$$

$$= \frac{\underset{(1\times p)}{\mathbf{y}'} \cdot \underset{(p\times p)}{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}} \cdot \underset{(p\times 1)}{\mathbf{y}}}{\underset{(1\times p)}{\mathbf{y}'} \cdot \underset{(p\times 1)}{\mathbf{y}}}$$

$$= \frac{\underset{(1\times p)}{\mathbf{y}'} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times 1)}{\mathbf{y}}}{\underset{(1\times p)}{\mathbf{y}'} \cdot \underset{(p\times 1)}{\mathbf{y}}}$$

$$= \frac{\underset{(1\times p)}{\left[y_1, y_2, \dots, y_p\right]} \cdot \underset{(p\times p)}{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}} \cdot \underset{(p\times 1)}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}}}{\underset{(1\times p)}{\left[y_1, y_2, \dots, y_p\right]} \cdot \underset{(p\times 1)}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}}}$$

$$= \frac{\underset{(1\times p)}{\left[y_1, y_2, \dots, y_p\right]} \cdot \underset{(p\times 1)}{\begin{bmatrix} \lambda_1 y_1 \\ \lambda_2 y_2 \\ \vdots \\ \lambda_p y_p \end{bmatrix}}}{\sum_{i=1}^{p} y_i^2}$$

$$= \frac{\sum_{i=1}^{p} \lambda_i y_i^2}{\sum_{i=1}^{p} y_i^2}$$

$$\leq \lambda_1 \cdot \frac{\sum_{i=1}^{p} y_i^2}{\sum_{i=1}^{p} y_i^2}$$

$$= \lambda_1$$

Setting,

$$\underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times 1)}{\mathbf{e}_1} = \underset{(p\times 1)}{\begin{bmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{p1} \end{bmatrix}}$$

gives

$$\underset{(p\times 1)}{\mathbf{y}}$$

$$= \underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times 1)}{\mathbf{e}_1}$$

$$= \underset{(p\times p)}{\begin{bmatrix} e_{11} & e_{21} & \cdots & e_{p1} \\ e_{12} & e_{22} & \cdots & e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{2p} & \cdots & e_{pp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{p1} \end{bmatrix}}$$

$$= \underset{(p\times 1)}{\begin{bmatrix} \mathbf{e}_1'\mathbf{e}_1 \\ \mathbf{e}_2'\mathbf{e}_1 \\ \vdots \\ \mathbf{e}_p'\mathbf{e}_1 \end{bmatrix}}$$

$$= \underset{(p\times 1)}{\begin{bmatrix} 1_1 \\ 0_2 \\ \vdots \\ 0_p \end{bmatrix}}, \{\text{orthogonality of eigenvectors}\}$$

$$= \underset{(p\times 1)}{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}$$

That is,

$$\mathbf{e}_k'\mathbf{e}_1 = \begin{cases} 1, & k = 1 \\ 0, & k = 0 \end{cases}$$

For this choice of $\underset{(p\times 1)}{\mathbf{x}}$, we have $\underset{(p\times 1)}{\mathbf{y}} = \underset{(p\times 1)}{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}} \Rightarrow$

$$= \frac{\underset{(1\times p)}{\mathbf{y}'} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times 1)}{\mathbf{y}}}{\underset{(1\times p)}{\mathbf{y}'} \cdot \underset{(p\times 1)}{\mathbf{y}}}$$

$$= \frac{\underset{(1\times p)}{[1_1, 0_2, \dots, 0_p]} \cdot \underset{(p\times p)}{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}} \cdot \underset{(p\times 1)}{\begin{bmatrix} 1_1 \\ 0_2 \\ \vdots \\ 0_p \end{bmatrix}}}{\underset{(1\times p)}{[1_1, 0_2, \dots, 0_p]} \cdot \underset{(p\times 1)}{\begin{bmatrix} 1_1 \\ 0_2 \\ \vdots \\ 0_p \end{bmatrix}}}$$

$$= \frac{\underset{(1\times p)}{[1_1, 0_2, \dots, 0_p]} \cdot \underset{(p\times 1)}{\begin{bmatrix} \lambda_1 \\ 0_2 \\ \vdots \\ 0_p \end{bmatrix}}}{1} = \frac{\lambda_1}{1} = \lambda_1$$

A similar argument produces the second part.

Now,

$$\underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times 1)}{\mathbf{y}} = y_1 \cdot \underset{(p\times 1)}{\mathbf{e}_1} + y_2 \cdot \underset{(p\times 1)}{\mathbf{e}_2} + \cdots + y_p \cdot \underset{(p\times 1)}{\mathbf{e}_p},$$

so $\underset{(p\times 1)}{\mathbf{x}} \perp \underset{(p\times 1)}{\mathbf{e}_1}, \underset{(p\times 1)}{\mathbf{e}_2}, \dots, \underset{(p\times 1)}{\mathbf{e}_k}$ implies

$$0 = \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times 1)}{\mathbf{x}} = y_1 \cdot \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times 1)}{\mathbf{e}_1} + \cdots + y_p \cdot \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times 1)}{\mathbf{e}_p} = y_i, \qquad i \le k$$

Therefore, for $\underset{(p\times 1)}{\mathbf{x}}$ perpendicular to the first k eigenvectors $\underset{(p\times 1)}{\mathbf{e}_i}$, the left-hand side

of the inequality in becomes

$$\frac{\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times 1)}{\mathbf{x}}} = \frac{\sum_{i=k+1}^{p} \lambda_i y_i^2}{\sum_{i=k+1}^{p} y_i^2}$$

Taking $y_{k+1} = 1, y_{k+2} = \cdots = y_p = 0$ gives the asserted maximum. ■

For a fixed $\underset{(p\times 1)}{\mathbf{x}_0} \neq \underset{(p\times 1)}{\mathbf{0}}$,

$$\frac{\underset{(1\times p)}{\mathbf{x}_0'} \cdot \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times 1)}{\mathbf{x}_0}}{\underset{(1\times p)}{\mathbf{x}_0'} \cdot \underset{(p\times 1)}{\mathbf{x}_0}}$$

has the same value as

$$\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times 1)}{\mathbf{x}},$$

where

$$\underset{(1\times p)}{\mathbf{x}'} = \frac{\underset{(1\times p)}{\mathbf{x}_0'}}{\sqrt{\underset{(1\times p)}{\mathbf{x}_0'} \cdot \underset{(p\times 1)}{\mathbf{x}_0}}} = \frac{\underset{(1\times p)}{\mathbf{x}_0'}}{\underset{(1\times 1)}{L_{\mathbf{x}_0}}}$$

is of unit length. Consequently,

$$\max_{\underset{(p\times 1)}{\mathbf{x}} \neq \underset{(p\times 1)}{\mathbf{0}}} \frac{\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times 1)}{\mathbf{x}}}{\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times 1)}{\mathbf{x}}} = \lambda_1 \qquad \left(attained\ when \underset{(p\times 1)}{\mathbf{x}} = \underset{(p\times 1)}{\mathbf{e}_1} \right)$$

says that the largest eigenvalue, λ_1, is the maximum value of the quadratic form

$$\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times 1)}{\mathbf{x}}$$

for all points $\underset{(p\times 1)}{\mathbf{x}}$ whose distance from the origin is unity. Similarly, λ_p is the smallest value of the quadratic form for all points $\underset{(p\times 1)}{\mathbf{x}}$ one unit from the origin. The largest and smallest eigenvalues thus represent extreme values of

$$\underset{(1\times p)}{\mathbf{x}'} \cdot \underset{(p\times p)}{\mathbf{B}} \cdot \underset{(p\times 1)}{\mathbf{x}}$$

for points on the unit sphere. The "intermediate" eigenvalues of the $p \times p$ positive definite matrix $\underset{(p\times p)}{\mathbf{B}}$ also have an interpretation as extreme values when $\underset{(p\times 1)}{\mathbf{x}}$ is further restricted to be perpendicular to the earlier choices [3, pp. 80-81].

Chapter 3

Multivariate Population Theory

3.1 Population Random Matrix

Definition 3.1.1 (Population Random Matrix **X**). *A* **population random matrix** $\underset{(n\times p)}{\mathbf{X}}$ *for continuous variables is a matrix whose elements are population continuous random variables. Specifically, let* $\underset{(n\times p)}{\mathbf{X}} = \{X_{ij}\}$ *be an* $n \times p$ *population random matrix*

$$\underset{(n\times p)}{\mathbf{X}} = \underset{(n\times p)}{\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1j} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2j} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{i1} & X_{i2} & \cdots & X_{ij} & \cdots & X_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nj} & \cdots & X_{np} \end{bmatrix}} = \underset{(n\times p)}{\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}}$$

for $i = 1,2, \ldots, n$ *rows and* $j = 1,2, \ldots, p$ *columns* [3, p. 66].

3.2 Population Random Vector, Mean Vector, Variance-Covariance Matrix, and Correlation Matrix

3.2.1 Population Random Vector

Definition 3.2.1 (Population Random Vector **X**). *A* **population random vector** $\underset{(p\times 1)}{\mathbf{X}}$ *for continuous random variables is a vector whose elements are population continuous random variables from a* $p-variate$ *population. Specifically, let* $\underset{(p\times 1)}{\mathbf{X}} = \{X_i\}$ *be a* $p \times 1$ *population random vector*

$$\underset{(p\times 1)}{\mathbf{X}} = \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}}$$

[3, p. 68].

3.2.2 Probability Density Functions

Definition 3.2.2 (Joint Probability Density Function). *The collective behavior of the* p *continuous random variables* $X_1, X_2, \dots, X_p$ *or, equivalently, the population random vector* $\underset{(p\times 1)}{\mathbf{X}}$, *is described by a* **joint probability density function (pdf)**

$$f\left(\underset{(p\times 1)}{\mathbf{x}}\right) = f_{12\cdots p}(x_1, x_2, \dots, x_p)$$

[3, p. 68] *where* $x_i \in \mathbb{R}, i = 1,2,\dots,p$. *Satisfying constraints,*

(**a**) $f_{12\cdots p}(x_1, x_2, \dots, x_p) \geq 0$

(**b**) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{12\cdots p}(x_1, x_2, \dots, x_p)dx_1 dx_2 \cdots dx_p = 1$

Definition 3.2.3 (Univariate Marginal Probability Density Function). *Each element of* $\underset{(p\times 1)}{\mathbf{X}}$ *is a population random variable with its own* **univariate marginal pdf** *defined as* $f_i(x_i)$. *Specifically,*

$$f_i(x_i) = \begin{cases} \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{12\cdots p}(x_1, x_2, \dots, x_p)dx_1 dx_2 \cdots dx_{i-1}dx_{i+1}\cdots dx_p \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

[3, p. 68] *for* $x_i \in \mathbb{R}, i = 1,2,\dots,p$. *Satisfying constraints,*

(**a**) $f_i(x_i) \geq 0$

(**b**) $\displaystyle\int_{-\infty}^{\infty} f_i(x_i)dx_i = 1$

Definition 3.2.4 (Bivariate Marginal Probability Density Function). *Each pair of elements of* $\underset{(p\times 1)}{\mathbf{X}}$ *is a bivariate population random vector* (X_i, X_k) *with a* **bivariate (joint) marginal pdf** *defined as* $f_{ik}(x_i, x_k)$*. Specifically,*

$$f_{ik}(x_i, x_k) = \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{12\cdots p}(x_1, x_2, \dots, x_p)dx_1 dx_2 \cdots dx_{i-1}dx_{i+1}\cdots dx_{k-1}dx_{k+1}\cdots dx_p & \\ 0 & \text{otherwise} \end{cases}$$

[3, p. 68] *for* $(x_i, x_k) \in \mathbb{R}$, $i, k = 1,2, \dots, p, i \neq k$*. Satisfying constraints,*

(**a**) $f_{ik}(x_i, x_k) \geq 0$

(**b**) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{ik}(x_i, x_k)dx_i dx_k = 1$

3.2.3 Population Parameters

Definition 3.2.5 (Univariate Marginal Population Mean). *The* **univariate marginal population means** μ_i *are defined as* $\mu_i = E(X_i)$ *with pdf* $f_i(x_i)$*. Specifically, if they exist (finite)*

$$\mu_i = E(X_i) = \int_{-\infty}^{\infty} x_i f_i(x_i)dx_i$$

for $i = 1,2, \dots, p$ *where* $-\infty < \mu_i < \infty$ [3, p. 68].

Definition 3.2.6 (Univariate Marginal Population Variance). *The* **univariate marginal population variances** σ_{ii} *are defined as* $\sigma_{ii} = \mathrm{var}(X_i) = E(X_i - \mu_i)^2$ *with pdf* $f_i(x_i)$. *Specifically, if they exist*

$$\sigma_{ii} = \mathrm{var}(X_i) = E(X_i - \mu_i)^2 = \int_{-\infty}^{\infty} (x_i - \mu_i)^2 f_i(x_i) dx_i$$

for $i = 1,2, \dots, p$ *where* $0 < \sigma_{ii} < \infty$. *The* **univariate marginal population standard deviation** *is the square-root of the variance* $\sqrt{\sigma_{ii}}$ [3, p. 68].

Definition 3.2.7 (Bivariate Marginal Population Covariance). *The* **bivariate marginal population covariances** σ_{ik} *are defined as* $\sigma_{ik} = \mathrm{cov}(X_i, X_k) = E(X_i - \mu_i)(X_k - \mu_k)$ *with pdf* $f_{ik}(x_i, x_k)$. *Specifically, if they exist*

$$\sigma_{ik} = \mathrm{cov}(X_i, X_k) = E(X_i - \mu_i)(X_k - \mu_k)$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_i - \mu_i)(x_k - \mu_k) f_{ik}(x_i, x_k) dx_i dx_k$$

for $i, k = 1,2, \dots, p$ *where* $-\infty < \sigma_{ik} < \infty$ [3, p. 68].

Note that $\sigma_{ik} = \sigma_{ki}$ and when $i = k$ the bivariate marginal population covariance becomes the univariate marginal population variance σ_{ii}.

Definition 3.2.8 (Bivariate Marginal Population Correlation). *The* **bivariate marginal population correlations** ρ_{ik} *are defined as* $\rho_{ik} = \mathrm{corr}(X_i, X_k) = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}$. *Specifically, if they exist*

$$\rho_{ik} = \mathrm{corr}(X_i, X_k) = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}$$

$$= \frac{E(X_i - \mu_i)(X_k - \mu_k)}{\sqrt{E(X_i - \mu_i)^2}\sqrt{E(X_k - \mu_k)^2}}$$

$$= \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i - \mu_i)(x_k - \mu_k)f_{ik}(x_i, x_k)dx_i dx_k}{\sqrt{\int_{-\infty}^{\infty}(x_i - \mu_i)^2 f_i(x_i)dx_i}\sqrt{\int_{-\infty}^{\infty}(x_k - \mu_k)^2 f_k(x_k)dx_k}}$$

for $i, k = 1,2, \dots, p$ *where* $-1 \leq \rho_{ik} \leq 1$ [3, p. 71].

Note that $\rho_{ik} = \rho_{ki}$ and when $i = k$ the bivariate marginal population correlation becomes $\rho_{ii} = \frac{\sigma_{ii}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{ii}}} = \frac{\sigma_{ii}}{\sigma_{ii}} = 1$.

3.2.4 Independent Random Variables

Definition 3.2.9 (Statistically Independent). *If the bivariate marginal pdf* $f_{ik}(x_i, x_k)$ *for continuous random variables* (X_i, X_k)*, can be written as the product of the corresponding univariate marginal pdf's* $f_i(x_i)$, $f_k(x_k)$ *so that*

$$f_{ik}(x_i, x_k) \equiv f_i(x_i)f_k(x_k)$$

then X_i *and* X_k *are said to be* **statistically independent**.

Furthermore, if, (X_i, X_k) are statistical independent, then $\sigma_{ik} = 0$ and $\rho_{ik} = 0$ [3, pp. 69, 71].

Definition 3.2.10 (Mutually Statistically Independent). *The* p *population continuous random variables* $(X_1, X_2, \dots, X_p)$ *are* **mutually statistically independent** *if their joint pdf can be factored as a product of their univariate marginal pdf's*

$$f_{12\cdots p}(x_1, x_2, \dots, x_p) \equiv f_1(x_1)f_2(x_2)\cdots f_p(x_p)$$

[3, p. 69].

In addition, if, $(X_1, X_2, \dots, X_p)$ are mutually statistically independent, then every subset of continuous population random variables ≥ 2 are also mutually statistically independent.

3.2.5 Population Mean Vector

Definition 3.2.11 (Population Mean Vector for **X**). *The* **population mean vector for** $\underset{(p\times1)}{\mathbf{X}}$ *or expected value of a population random vector is a random vector consisting of the univariate marginal expectations of each of its elements. Then, if these expectations exist, the population mean vector for* $\underset{(p\times1)}{\mathbf{X}}$*, denoted by* $\underset{(p\times1)}{\boldsymbol{\mu}_{\mathbf{X}}} = \underset{(p\times1)}{E(\mathbf{X})}$*, is the* $p \times 1$ *vector*

$$\underset{(p\times1)}{\boldsymbol{\mu}_{\mathbf{X}}} = \underset{(p\times1)}{E(\mathbf{X})} = \underset{(p\times1)}{\begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix}} = \underset{(p\times1)}{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}}$$

where $-\infty < \mu_i < \infty$*, for* $i = 1,2, \dots, p$ [3, p. 69].

3.2.6 Population Variance-Covariance Matrix

Theorem 3.2.1 (Population Variance-Covariance Matrix for **X**). *The* **population variance-covariance matrix for** $\underset{(p\times 1)}{\mathbf{X}}$ *is a symmetric matrix containing the* p *univariate marginal population variances* σ_{ii} *and the* $p(p-1)/2$ *distinct bivariate marginal population covariances* σ_{ik} $(i<k)$. *Then, if these variances and covariances exist, the* $p\times p$ *population variance-covariance matrix for* $\underset{(p\times 1)}{\mathbf{X}}$ *is given by*

$$\underset{(p\times p)}{\Sigma_{\mathbf{X}}} = \underset{(p\times p)}{\mathrm{Cov}(\mathbf{X})} = E\underset{(p\times 1)}{(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})}\underset{(1\times p)}{(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})'}$$

where $\underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} = \underset{(p\times 1)}{E(\mathbf{X})}$ *is the population mean vector.*

Proof. Use linearity of the operator E, Definition 2.1.2, 2.1.11, 2.2.5, 3.2.6., and 3.2.7.

$$\underset{(p\times p)}{\Sigma_{\mathbf{X}}}$$

$$= \underset{(p\times p)}{\mathrm{Cov}(\mathbf{X})}$$

$$= E\underset{(p\times 1)}{(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})}\underset{(1\times p)}{(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})'}$$

$$= E\left(\underset{(p\times 1)}{\begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \\ X_p-\mu_p \end{bmatrix}} \cdot \underset{(1\times p)}{[X_1-\mu_1, X_2-\mu_2, \dots, X_p-\mu_p]}\right)$$

$$= E\left(\underset{(p\times p)}{\begin{bmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \cdots & (X_1-\mu_1)(X_p-\mu_p) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & \cdots & (X_2-\mu_2)(X_p-\mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p-\mu_p)(X_1-\mu_1) & (X_p-\mu_p)(X_2-\mu_2) & \cdots & (X_p-\mu_p)^2 \end{bmatrix}}\right)$$

$$= E\underset{(p\times p)}{\begin{bmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \cdots & (X_1-\mu_1)(X_p-\mu_p) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & \cdots & (X_2-\mu_2)(X_p-\mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p-\mu_p)(X_1-\mu_1) & (X_p-\mu_p)(X_2-\mu_2) & \cdots & (X_p-\mu_p)^2 \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} E(X_1-\mu_1)^2 & E(X_1-\mu_1)(X_2-\mu_2) & \cdots & E(X_1-\mu_1)(X_p-\mu_p) \\ E(X_2-\mu_2)(X_1-\mu_1) & E(X_2-\mu_2)^2 & \cdots & E(X_2-\mu_2)(X_p-\mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p-\mu_p)(X_1-\mu_1) & E(X_p-\mu_p)(X_2-\mu_2) & \cdots & E(X_p-\mu_p)^2 \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}}$$

[3, pp. 69-70] ■

3.2.7 Population Standard Deviation Matrix

Definition 3.2.12 (Population Standard Deviation Matrix for **X**). *The* **population standard deviation matrix for** $\underset{(p\times 1)}{\mathbf{X}}$ *is a diagonal matrix containing the* p *univariate marginal population standard deviations* $\sqrt{\sigma_{ii}}$ *along the main diagonal. Then, if these standard deviations exist, the population standard deviation matrix for* $\underset{(p\times 1)}{\mathbf{X}}$ *is denoted by* $\underset{(p\times p)}{\mathbf{V}^{1/2}}$, *is the* $p \times p$ *matrix*

$$\underset{(p\times p)}{\mathbf{V}^{1/2}} = \underset{(p\times p)}{\begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}}$$

with inverse

$$\underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}} = \underset{(p\times p)}{\mathbf{V}^{-1/2}} = \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{bmatrix}}$$

[3, pp. 59, 72].

3.2.8 Population Correlation Matrix

Theorem 3.2.2 (Population Correlation Matrix for $\mathbf{X}$). *The* **population correlation matrix for** $\underset{(p\times 1)}{\mathbf{X}}$ *is a symmetric matrix containing the* p *bivariate marginal population correlations* $\rho_{ii} = 1$ *along the main diagonal and the* $p(p-1)/2$ *distinct bivariate marginal population correlations* ρ_{ik} $(i < k)$. *Then, if these correlations exist, the* $p \times p$ *population correlation matrix for* $\underset{(p\times 1)}{\mathbf{X}}$ *is given by*

$$\underset{(p\times p)}{\boldsymbol{\rho}} = \underset{(p\times p)}{\mathrm{Corr}(\mathbf{X})} = \underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}}$$

where $\underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}}$ *is the inverse population standard deviation matrix and* $\underset{(p\times p)}{\sum_{\mathbf{X}}}$ *is the population variance-covariance matrix* [3, p. 72].

Proof. Use Definition 2.2.5, 3.2.6. and 3.2.7.

$$\underset{(p\times p)}{\boldsymbol{\rho}} = \underset{(p\times p)}{\mathrm{Corr}(\mathbf{X})}$$

$$= \underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}}$$

$$= \underset{(p\times p)}{\mathbf{V}^{-1/2}} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times p)}{\mathbf{V}^{-1/2}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{bmatrix}} \cdot \underset{(p\times p)}{\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}} \cdot \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{bmatrix}}$$

$$\underset{(p\times p)}{\begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}}} & \frac{\sigma_{p2}}{\sqrt{\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}} \end{bmatrix}} \cdot \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{11}}} & \frac{\sigma_{p2}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pp} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}}$$

Thus, $\underset{(p\times p)}{\boldsymbol{\rho}}$ can be obtained from $\underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}}$ and $\sum_{\mathbf{X}}$ ■

Corollary 3.2.1. *Let* $\underset{(p\times p)}{\mathbf{V}^{1/2}}$ *be the population standard deviation matrix and* $\underset{(p\times p)}{\boldsymbol{\rho}}$ *be the population correlation matrix. Then* $\underset{(p\times p)}{\sum_{\mathbf{X}}}$ *the population variance-covariance matrix can be obtained. That is,*

$$\underset{(p\times p)}{\boldsymbol{\rho}} = \underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}}$$

$$\underset{(p\times p)}{\mathbf{V}^{1/2}} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times p)}{\mathbf{V}^{1/2}} = \underset{(p\times p)}{\mathbf{V}^{1/2}} \cdot \underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}} \cdot \underset{(p\times p)}{\mathbf{V}^{1/2}}$$

$$\underset{(p\times p)}{\mathbf{V}^{1/2}} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times p)}{\mathbf{V}^{1/2}} = \underset{(p\times p)}{\mathbf{I}} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times p)}{\mathbf{I}}, \left\{\underset{(p\times p)}{(\mathbf{V}^{1/2})^{-1}} \text{ inverse of } \underset{(p\times p)}{\mathbf{V}^{1/2}}\right\}$$

$$\underset{(p\times p)}{\sum_{\mathbf{X}}} = \underset{(p\times p)}{\mathbf{V}^{1/2}} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times p)}{\mathbf{V}^{1/2}}$$

[3, p. 72].

3.3 Population Mean Vector and Variance-Covariance Matrix for Linear Combinations of Continuous Random Variables

3.3.1 Linear Combination

Definition 3.3.1 (Linear Combination of **X**). *Let* $\underset{(p\times 1)}{\mathbf{c}}$ *be a* $p \times 1$ *vector of constants defined as*

$$\underset{(p\times 1)}{\mathbf{c}} = \underset{(p\times 1)}{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}}$$

and let $\underset{(p\times 1)}{\mathbf{X}}$ *be a* $p \times 1$ *population random vector of continuous random variables*

$$\underset{(p\times 1)}{\mathbf{X}} = \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}}$$

$i = 1,2, \dots, p$. *Then a* **linear combination of** $\underset{(p\times 1)}{\mathbf{X}}$, *is given by the inner product*

$$\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}} = \underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = c_1 X_1 + c_2 X_2 + \cdots + c_p X_p$$

[3, p. 76].

3.3.2 Population Parameters for Linear Combinations

Theorem 3.3.1 (Mean of a Linear Combination of $\mathbf{X}$). *Suppose a linear combination* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *is given by Definition 3.3.1 and a population mean vector* $\underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} = \underset{(p\times 1)}{E(\mathbf{X})}$ *is given by Definition 3.2.11. Then the* **expected value** *or* **mean of a linear combination of** $\underset{(p\times 1)}{\mathbf{X}}$, *is given by*

$$E\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}}$$

Proof. Using linearity of E and Definition 3.2.5.

$E\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right)$

$$= E\left(\underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} \right)$$

$$= E(c_1X_1 + c_2X_2 + \cdots + c_pX_p)$$

$$= c_1E(X_1) + c_2E(X_2) + \cdots + c_pE(X_p)$$

$$= c_1\mu_1 + c_2\mu_2 + \cdots + c_p\mu_p$$

$$= \underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \underset{(p\times 1)}{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}}$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} \ \blacksquare$$

Theorem 3.3.2 (Variance of a Linear Combination of **X**). *Suppose a linear combination* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *is given by Definition 3.3.1 and a population variance-covariance* $\underset{(p\times p)}{\Sigma_{\mathbf{X}}} = \underset{(p\times p)}{\mathrm{Cov}(\mathbf{X})}$ *is given by Theorem 3.2.1. Then the* **variance of a linear combination of** $\underset{(p\times 1)}{\mathbf{X}}$, *is given by*

$$\mathrm{var}\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}} = \sum_{i}^{p} \sum_{k=1}^{p} c_i c_k \sigma_{ik}$$

$$= \sum_{i=1}^{p} c_i^2 \sigma_{ii} + \underbrace{\sum \sum c_i c_k \sigma_{ik}}_{i \neq k} = \sum_{i=1}^{p} c_i^2 \sigma_{ii} + 2 \sum_{i<k} c_i c_k \sigma_{ik}$$

Proof. Using properties of variance and covariance.

$$\operatorname{var}\left(\underset{(1\times p)}{\mathbf{c}'}\cdot\underset{(p\times 1)}{\mathbf{X}}\right)$$

$$=\operatorname{var}\left(\underset{(1\times p)}{\left[c_1,c_2,\dots,c_p\right]}\underset{(p\times 1)}{\begin{bmatrix}X_1\\X_2\\\vdots\\X_p\end{bmatrix}}\right)$$

$$=\operatorname{var}\left(c_1X_1+c_2X_2+\cdots+c_pX_p\right)$$

$$=\operatorname{cov}\left(c_1X_1+c_2X_2+\cdots+c_pX_p,c_1X_1+c_2X_2+\cdots+c_pX_p\right)$$

$$=c_1^2\operatorname{var}(X_1)+c_1c_2\operatorname{cov}(X_1,X_2)+\cdots+c_1c_p\operatorname{cov}\left(X_1,X_p\right)$$

$$+c_2c_1\operatorname{cov}(X_2,X_1)+c_2^2\operatorname{var}(X_2)+\cdots+c_2c_p\operatorname{cov}\left(X_2,X_p\right)$$

$$+\cdots+$$

$$c_pc_1\operatorname{cov}\left(X_p,X_1\right)+c_pc_2\operatorname{cov}\left(X_p,X_2\right)+\cdots+c_p^2\operatorname{var}\left(X_p\right)$$

$$=c_1^2\operatorname{var}(X_1)+c_2^2\operatorname{var}(X_2)+\cdots+c_p^2\operatorname{var}\left(X_p\right)$$

$$+2c_1c_2\operatorname{cov}(X_1,X_2)+\cdots+2c_{p-1}c_p\operatorname{cov}\left(X_{p-1},X_p\right)$$

$$=c_1^2\sigma_{11}+c_2^2\sigma_{22}+\cdots+c_p^2\sigma_{pp}$$

$$+2c_1c_2\sigma_{12}+\cdots+2c_{p-1}c_p\sigma_{(p-1)(p)}$$

$$=\sum_{i=1}^{p}\sum_{k=1}^{p}c_ic_k\sigma_{ik}=\sum_{i=1}^{p}c_i^2\sigma_{ii}+\underbrace{\sum\sum c_ic_k\sigma_{ik}}_{i\neq k}$$

$$=\sum_{i=1}^{p}c_i^2\sigma_{ii}+2\sum_{i<k}c_ic_k\sigma_{ik}$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$$= \underset{(1\times p)}{\left[c_1, c_2, \dots, c_p\right]} \underset{(p\times p)}{\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}}$$

$$= \underset{(1\times p)}{\left[c_1, c_2, \dots, c_p\right]} \underset{(p\times 1)}{\begin{bmatrix} c_1\sigma_{11} + c_2\sigma_{12} + \cdots + c_p\sigma_{1p} \\ c_1\sigma_{21} + c_2\sigma_{22} + \cdots + c_p\sigma_{2p} \\ \vdots \\ c_1\sigma_{p1} + c_2\sigma_{p2} + \cdots + c_p\sigma_{pp} \end{bmatrix}}$$

$$= c_1(c_1\sigma_{11} + c_2\sigma_{12} + \cdots + c_p\sigma_{1p})$$

$$+c_2(c_2c_1\sigma_{21} + c_2\sigma_{22} + \cdots + c_p\sigma_{2p})$$

$$+ \cdots +$$

$$c_p(c_1\sigma_{p1} + c_2\sigma_{p2} + \cdots + c_p\sigma_{pp})$$

$$= c_1^2\sigma_{11} + c_1c_2\sigma_{12} + \cdots + c_1c_p\sigma_{1p}$$

$$+c_2c_1\sigma_{21} + c_2^2\sigma_{22} + \cdots + c_2c_p\sigma_{2p}$$

$$+ \cdots +$$

$$c_pc_1\sigma_{p1} + c_pc_2\sigma_{p2} + \cdots + c_p^2\sigma_{pp}$$

$$= c_1^2\sigma_{11} + c_2^2\sigma_{22} + \cdots + c_p^2\sigma_{pp}$$

$$+2c_1c_2\sigma_{12} + \cdots + 2c_{p-1}c_p\sigma_{(p-1)(p)}$$

$$= \sum_{i=1}^{p}\sum_{k=1}^{p} c_ic_k\sigma_{ik} = \sum_{i=1}^{p} c_i^2\sigma_{ii} + \underbrace{\sum\sum c_ic_k\sigma_{ik}}_{i\neq k}$$

$$= \sum_{i=1}^{p} c_i^2\sigma_{ii} + 2\sum_{i<k} c_ic_k\sigma_{ik} \ \blacksquare$$

Theorem 3.3.3 (Covariance of Two Linear Combinations of **X**). *Suppose two linear combinations* $\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ and $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *are given following Definition 3.3.1 and a population variance-covariance* $\underset{(p\times p)}{\Sigma_{\mathbf{X}}} = \underset{(p\times p)}{\mathrm{Cov}(\mathbf{X})}$ *is given by Theorem 3.2.1.Then the* **covariance of two linear combinations of** $\underset{(p\times 1)}{\mathbf{X}}$, *is given by*

$$\mathrm{cov}\left(\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{X}}, \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$$= \sum_{i=1}^{p}\sum_{k=1}^{p} b_i c_k \sigma_{ik} = \sum_{i=1}^{p} b_i c_i \sigma_{ii} + \underbrace{\sum\sum b_i c_k \sigma_{ik}}_{i\neq k}$$

Proof. Using properties of variance and covariance.

$$\mathrm{cov}\left(\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{X}}, \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right)$$

$$= \mathrm{cov}\left(\underset{(1\times p)}{[b_1, b_2, \dots, b_p]} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}}, \underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}}\right)$$

$$= \mathrm{cov}(b_1X_1 + b_2X_2 + \cdots + b_pX_p, c_1X_1 + c_2X_2 + \cdots + c_pX_p)$$

$$= b_1c_1\text{var}(X_1) + b_1c_2\text{cov}(X_1, X_2) + \cdots + b_1c_p\text{cov}(X_1, X_p)$$

$$+b_2c_1\text{cov}(X_2, X_1) + b_2c_2\text{var}(X_2) + \cdots + b_2c_p\text{cov}(X_2, X_p)$$

$$+ \cdots +$$

$$b_pc_1\text{cov}(X_p, X_1) + b_pc_2\text{cov}(X_p, X_2) + \cdots + b_pc_p\text{var}(X_p)$$

$$= b_1c_1\sigma_{11} + b_2c_2\sigma_{22} + \cdots + b_pc_p\sigma_{pp} + b_1c_2\sigma_{12} + b_2c_1\sigma_{21}$$

$$+ \cdots + b_{p-1}c_p\sigma_{(p-1)(p)} + b_pc_{p-1}\sigma_{(p)(p-1)}$$

$$= \sum_{i=1}^{p}\sum_{k=1}^{p} b_ic_k\sigma_{ik}$$

$$= \sum_{i=1}^{p} b_ic_i\sigma_{ii} + \underbrace{\sum\sum b_ic_k\sigma_{ik}}_{i\neq k}$$

$$= \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times p)}{\sum\mathbf{x}} \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$$= \underset{(1\times p)}{[b_1, b_2, \dots, b_p]} \underset{(p\times p)}{\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}}$$

$$= \underset{(1\times p)}{[b_1, b_2, \dots, b_p]} \underset{(p\times 1)}{\begin{bmatrix} c_1\sigma_{11} + c_2\sigma_{12} + \cdots + c_p\sigma_{1p} \\ c_1\sigma_{21} + c_2\sigma_{22} + \cdots + c_p\sigma_{2p} \\ \vdots \\ c_1\sigma_{p1} + c_2\sigma_{p2} + \cdots + c_p\sigma_{pp} \end{bmatrix}}$$

$$= b_1\left(c_1\sigma_{11} + c_2\sigma_{12} + \cdots + c_p\sigma_{1p}\right)$$

$$+b_2(c_1\sigma_{21} + c_2\sigma_{22} + \cdots + c_p\sigma_{2p})$$

$$+\cdots+$$

$$b_p(c_1\sigma_{p1} + c_2\sigma_{p2} + \cdots + c_p\sigma_{pp})$$

$$= b_1c_1\sigma_{11} + b_1c_2\sigma_{12} + \cdots + b_1c_p\sigma_{1p}$$

$$+b_2c_1\sigma_{21} + b_2c_2\sigma_{22} + \cdots + b_2c_p\sigma_{2p}$$

$$+\cdots+$$

$$b_pc_1\sigma_{p1} + b_pc_2\sigma_{p2} + \cdots + b_pc_p\sigma_{pp}$$

$$= b_1c_1\sigma_{11} + b_2c_2\sigma_{22} + \cdots + b_pc_p\sigma_{pp} + b_1c_2\sigma_{12} + b_2c_1\sigma_{21}$$

$$+\cdots + b_{p-1}c_p\sigma_{(p-1)(p)} + b_pc_{p-1}\sigma_{(p)(p-1)}$$

$$= \sum_{i=1}^{p}\sum_{k=1}^{p} b_ic_k\sigma_{ik} = \sum_{i=1}^{p} b_ic_i\sigma_{ii} + \underbrace{\sum\sum b_ic_k\sigma_{ik}}_{i\neq k} \blacksquare$$

3.3.3 q Linear Combinations

Definition 3.3.2 (q Linear Combinations of $\mathbf{X}$). *Consider* $\underset{(q\times p)}{\mathbf{C}}$ *a matrix of real constants and the* ***q*** **linear combinations of** $\underset{(p\times 1)}{\mathbf{X}}$, Y_i,

$$Y_1 = \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}} = c_{11}X_1 + c_{12}X_2 + \cdots + c_{1p}X_p$$

$$Y_2 = \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}} = c_{21}X_1 + c_{22}X_2 + \cdots + c_{2p}X_p$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$Y_q = \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\mathbf{X}} = c_{q1}X_1 + c_{q2}X_2 + \cdots + c_{qp}X_p$$

or in matrix notation,

$$\underset{(q\times 1)}{\mathbf{Y}} = \underset{(q\times 1)}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{bmatrix}} = \underset{(q\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\mathbf{X}} \end{bmatrix}} = \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{X}}$$

[3, p. 76].

3.3.4 Population Mean Vector for q Linear Combinations

Theorem 3.3.4 (Population Mean Vector for q Linear Combinations of $\mathbf{X}$). *Suppose* q *linear combinations* $Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *are given by Definition 3.3.2 and a population mean vector* $\underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} = \underset{(p\times 1)}{E(\mathbf{X})}$ *is given by Definition 3.2.11.Then the* **population mean vector for q linear combinations of** $\underset{(p\times 1)}{\mathbf{X}}$, $\underset{(p\times 1)}{\mathbf{Y}}$, *is given by*

$$\underset{(q\times 1)}{\boldsymbol{\mu}_{\mathbf{Y}}} = \underset{(q\times 1)}{E(\mathbf{Y})} = E\left(\underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} = \underset{(q\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} \end{bmatrix}}$$

[3, p. 76].

Proof. Using the linearity of E and Definition 2.2.5.

$$\underset{(q\times 1)}{\boldsymbol{\mu}_{\mathbf{Y}}}$$

$$= \underset{(q\times 1)}{E(\mathbf{Y})}$$

$$= E\left(\underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{X}}\right)$$

$$= E\left(\underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}}\right)$$

$$= \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} E\left(\underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} \right)$$

$$= \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix}}$$

$$= \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}}$$

$$= \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}}$$

$$= \underset{(q\times 1)}{\begin{bmatrix} c_{11}\mu_1 + c_{12}\mu_2 + \cdots + c_{1p}\mu_p \\ c_{21}\mu_1 + c_{22}\mu_2 + \cdots + c_{2p}\mu_p \\ \vdots \\ c_{q1}\mu_1 + c_{q2}\mu_2 + \cdots + c_{qp}\mu_p \end{bmatrix}}$$

$$= \underset{(q\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} \end{bmatrix}} \blacksquare$$

Thus, the ith row of $\underset{(q\times 1)}{\mathbf{Y}}$ has population mean

$$\bar{Y}_i = E(Y_i) = E\left(\underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}}$$

for $i = 1,2,\dots,q$.

3.3.5 Population Variance-Covariance Matrix for q Linear Combinations

Theorem 3.3.5. (Population Variance-Covariance Matrix for q Linear Combinations of **X**). *Suppose* q *linear combinations* $Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *are given by Definition 3.3.2 and a population variance-covariance* $\underset{(p\times p)}{\sum_{\mathbf{X}}} = \underset{(p\times p)}{\mathrm{Cov}(\mathbf{X})}$ *is given by Theorem 3.2.1. Then the symmetric* **population variance-covariance matrix for q linear combinations of** $\underset{(p\times 1)}{\mathbf{X}}, \underset{(q\times 1)}{\mathbf{Y}}$, *is given by*

$$\underset{(q\times q)}{\sum_{\mathbf{Y}}} = \underset{(q\times q)}{\mathrm{Cov}(\mathbf{Y})} = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times q)}{\mathbf{C}'}$$

[3, p. 76].

Proof. Using Definition 2.2.5 for matrix multiplication and following Theorem 3.3.2 for computation of diagonal elements and Theorem 3.3.3 for computation of off-diagonal elements.

$$\underset{(q\times q)}{\sum_{\mathbf{Y}}}$$

$$= \underset{(q\times q)}{\mathrm{Cov}(\mathbf{Y})}$$

$$= \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} \cdot \underset{(p\times q)}{\mathbf{C}'}$$

$$= \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times p)}{\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}} \underset{(p\times q)}{\begin{bmatrix} c_{11} & c_{21} & \cdots & c_{q1} \\ c_{12} & c_{22} & \cdots & c_{q2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1p} & c_{2p} & \cdots & c_{qp} \end{bmatrix}}$$

$$= \underset{(q\times q)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \end{bmatrix}} \blacksquare$$

Thus, the ith row of $\underset{(q\times 1)}{\mathbf{Y}}$ has population variance

$$\text{var}(Y_i) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_i}$$

for $i = 1,2, \dots, q$.

And the ith row and kth row of $\underset{(q\times 1)}{\mathbf{Y}}$ have population covariance

$$\text{cov}(Y_i, Y_k) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_k} = \underset{(1\times p)}{\mathbf{c}_k'} \cdot \underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_i}$$

for $i, k = 1,2, \dots, q$.

3.4 Population Random Vector, Mean Vector, and Variance-Covariance Matrix for Standardized Continuous Random Variables

3.4.1 Population Random Vector for Standardized Continuous Random Variables

Definition 3.4.1 (Population Random Vector **Z**). *A population random vector for standardized continuous variables is a vector whose elements are standardized population continuous random variables from a* $p-variate$ *population. Each standardized continuous random variable is of the form*

$$Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}$$

for $i = 1,2,\dots,p$.

Specifically, let the **population random vector** $\underset{(p\times1)}{\mathbf{Z}} = \{Z_i\}$ *be* defined by

$$\underset{(p\times1)}{\mathbf{Z}} = \underset{(p\times p)}{\mathbf{V}^{-1/2}} \cdot \left(\underset{(p\times1)}{\mathbf{X}} - \underset{(p\times1)}{\boldsymbol{\mu}_{\mathbf{X}}}\right) = \underset{(p\times1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}} = \underset{(p\times1)}{\begin{bmatrix} \frac{X_1-\mu_1}{\sqrt{\sigma_{11}}} \\ \frac{X_2-\mu_2}{\sqrt{\sigma_{22}}} \\ \vdots \\ \frac{X_p-\mu_p}{\sqrt{\sigma_{pp}}} \end{bmatrix}}$$

where $\underset{(p\times1)}{\mathbf{X}}$ *is a population random vector defined in Definition 3.2.1.,* $\underset{(p\times1)}{\boldsymbol{\mu}_{\mathbf{X}}}$ *is a* *population mean vector defined in Definition 3.2.11., and* $\underset{(p\times p)}{\mathbf{V}^{-1/2}}$ *is an inverse population standard deviation matrix defined in Definition 3.2.12.*

$$\underset{(p\times 1)}{\mathbf{Z}} = \underset{(p\times p)}{\mathbf{V}^{-1/2}} \cdot \left(\underset{(p\times 1)}{\mathbf{X}} - \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}} \right)$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{bmatrix}} \cdot \left(\underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} - \underset{(p\times 1)}{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}} \right)$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{bmatrix}} \cdot \left(\underset{(p\times 1)}{\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix}} \right) = \underset{(p\times 1)}{\begin{bmatrix} \frac{X_1 - \mu_1}{\sqrt{\sigma_{11}}} \\ \frac{X_2 - \mu_2}{\sqrt{\sigma_{22}}} \\ \vdots \\ \frac{X_p - \mu_p}{\sqrt{\sigma_{pp}}} \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}}$$

[3, pp. 436-437].

3.4.2 Population Parameters for Standardized Continuous Random Variables

Theorem 3.4.1 (Univariate Marginal Population Mean for Z_i). *Suppose the univariate marginal population means* $\mu_i = E(X_i)$ *are given by Definition 3.2.5 and univariate marginal population standard deviations* $\sqrt{\sigma_{ii}}$ *are given by Definition 3.2.6. Then the* **univariate marginal population means for** Z_i *are given by*

$$\mu_{z,i} = E(Z_i) = E\left(\frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}\right) = 0$$

for $i = 1,2, \ldots, p$.

Proof. Using linearity of E.

$\mu_{z,i}$

$$= E(Z_i)$$

$$= E\left(\frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}\right)$$

$$= \frac{1}{\sqrt{\sigma_{ii}}} E(X_i - \mu_i)$$

$$= \frac{1}{\sqrt{\sigma_{ii}}} [E(X_i) - \mu_i]$$

$$= \frac{1}{\sqrt{\sigma_{ii}}} [\mu_i - \mu_i]$$

$$= 0 \blacksquare$$

Theorem 3.4.2 (Univariate Marginal Population Variance for Z_i). *Suppose the univariate marginal population means* $\mu_i = E(X_i)$ *are given by Definition 3.2.5 and univariate marginal population standard deviations* $\sqrt{\sigma_{ii}}$ *are given by Definition 3.2.6. Then the* **univariate marginal population variances for** Z_i *are given by*

$$\sigma_{z,ii} = E(Z_i - \mu_{z,i})^2 = \mathrm{var}(Z_i) = \mathrm{var}\left(\frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}\right) = 1$$

for $i = 1,2,\ldots,p$.

Proof. Using properties of variance and covariance.

$$\begin{aligned}
\sigma_{z,ii} & \\
&= E(Z_i - \mu_{z,i})^2 \\
&= \mathrm{var}(Z_i) \\
&= \mathrm{var}\left(\frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}\right) \\
&= \frac{1}{\sigma_{ii}} \mathrm{var}(X_i - \mu_i) \\
&= \frac{1}{\sigma_{ii}} \mathrm{cov}(X_i - \mu_i, X_i - \mu_i) \\
&= \frac{1}{\sigma_{ii}} [\mathrm{cov}(X_i, X_i) - \mathrm{cov}(X_i, \mu_i) - \mathrm{cov}(\mu_i, X_i) + \mathrm{cov}(\mu_i, \mu_i)] \\
&= \frac{1}{\sigma_{ii}} \mathrm{cov}(X_i, X_i) = \frac{1}{\sigma_{ii}} \mathrm{var}(X_i) \\
&= \frac{\sigma_{ii}}{\sigma_{ii}} = 1 \ \blacksquare
\end{aligned}$$

Theorem 3.4.3 (Bivariate Marginal Population Covariance for Z_i and Z_k). *Suppose the univariate marginal population means* $\mu_i = E(X_i)$ *are given by Definition 3.2.5 and univariate marginal population standard deviations* $\sqrt{\sigma_{ii}}$ *are given by Definition 3.2.6. Then the* **bivariate marginal population covariances for** Z_i **and** Z_k *are given by*

$$\sigma_{z,ik} = E\left(Z_i - \mu_{z,i}\right)\left(Z_k - \mu_{z,k}\right) = \operatorname{cov}(Z_i, Z_k) = \operatorname{cov}\left(\frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}, \frac{X_k - \mu_k}{\sqrt{\sigma_{kk}}}\right) = \rho_{ik}$$

for $i, k = 1,2, \ldots, p$.

Proof. Using properties of covariance and Definition 3.2.8.

$\sigma_{z,ik}$

$= E\left(Z_i - \mu_{z,i}\right)\left(Z_k - \mu_{z,k}\right)$

$= \operatorname{cov}(Z_i, Z_k)$

$$= \operatorname{cov}\left(\frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}, \frac{X_k - \mu_k}{\sqrt{\sigma_{kk}}}\right)$$

$$= \frac{1}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}} \operatorname{cov}(X_i - \mu_i, X_k - \mu_k)$$

$$= \frac{1}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}} \left[\operatorname{cov}(X_i, X_k) - \operatorname{cov}(X_i, \mu_k) - \operatorname{cov}(\mu_i, X_k) + \operatorname{cov}(\mu_i, \mu_k)\right]$$

$$= \frac{1}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}} \operatorname{cov}(X_i, X_k)$$

$$= \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}} = \operatorname{corr}(X_i, X_k) = \rho_{ik} \blacksquare$$

Thus, standardizing population continuous random variables turns bivariate marginal population covariances $\sigma_{z,ik}$ into bivariate marginal population correlations ρ_{ik}. That is, $\sigma_{z,ik} = \rho_{ik}$ for $i, k = 1,2, \dots, p$. If X_i, X_k are statistically independent, then $\sigma_{z,ik} = \rho_{ik} = 0$. Note $\sigma_{z,ik} = \sigma_{z,ki}$, and when $i = k$, $\sigma_{z,ii} = \rho_{ii} = 1$.

3.4.3 Population Mean Vector for Standardized Continuous Random Variables

Definition 3.4.2 (Population Mean Vector for $\mathbf{Z}$). *The* **population mean vector for** $\underset{(p\times 1)}{\mathbf{Z}}$ *or expected value of* $\underset{(p\times 1)}{\mathbf{Z}}$ *is a random vector consisting of the univariate marginal expectations of each of its standardized elements. Then the population mean vector* for $\underset{(p\times 1)}{\mathbf{Z}}$ *or expected value of* $\underset{(p\times 1)}{\mathbf{Z}}$ *denoted by* $\underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} = \underset{(p\times 1)}{E(\mathbf{Z})}$*, is the* $p \times 1$ *vector*

$$\underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} = \underset{(p\times 1)}{E(\mathbf{Z})} = \underset{(p\times 1)}{\begin{bmatrix} E(Z_1) \\ E(Z_2) \\ \vdots \\ E(Z_p) \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} \mu_{z,1} \\ \mu_{z,2} \\ \vdots \\ \mu_{z,p} \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} 0_{z,1} \\ 0_{z,2} \\ \vdots \\ 0_{z,p} \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}} = \underset{(p\times 1)}{\mathbf{0}}$$

Thus, the population mean vector for $\underset{(p\times 1)}{\mathbf{Z}}$ is the $\underset{(p\times 1)}{\mathbf{0}}$ − vector [3, p. 437].

3.4.4 Population Variance-Covariance Matrix for Standardized Continuous Random Variables

Theorem 3.4.4 (Population Variance-Covariance Matrix for $\mathbf{Z}$). *The* **population variance-covariance matrix for** $\underset{(p\times 1)}{\mathbf{Z}}$ *is a symmetric matrix containing the* p *univariate marginal population variances* $\sigma_{z,ii} = 1$ *and the* $p(p-1)/2$ *distinct bivariate marginal population covariances* $\sigma_{z,ik} = \rho_{ik}$ $(i < k)$*. Then, if these variances and covariances exist, the* $p \times p$ *population variance-covariance matrix for* $\underset{(p\times 1)}{\mathbf{Z}}$ *is given by*

$$\underset{(p\times p)}{\textstyle\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\mathrm{Cov}(\mathbf{Z})} = E\underset{(p\times 1)}{(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})}\underset{(1\times p)}{(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})'} = \underset{(p\times p)}{\boldsymbol{\rho}}$$

where $\underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} = \underset{(p\times 1)}{E(\mathbf{Z})} = \underset{(p\times 1)}{\mathbf{0}}$ *is the population mean vector for* $\underset{(p\times 1)}{\mathbf{Z}}$ [3, p. 437].

Proof. Use linearity of the operator E, Definition 2.1.2, 2.1.11, and 2.2.5, Theorem 3.4.2 and Theorem 3.4.3.

$$\underset{(p\times p)}{\textstyle\sum_{\mathbf{Z}}}$$

$$= \underset{(p\times p)}{\mathrm{Cov}(\mathbf{Z})}$$

$$= E\underset{(p\times 1)}{(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})}\underset{(1\times p)}{(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})'}$$

$$= E\left(\underset{(p\times 1)}{\begin{bmatrix} Z_1-\mu_{z,1} \\ Z_2-\mu_{z,2} \\ \vdots \\ Z_p-\mu_{z,p} \end{bmatrix}} \cdot \underset{(1\times p)}{[Z_1-\mu_{z,1}, Z_2-\mu_{z,2}, \dots, Z_p-\mu_{z,p}]}\right)$$

$$= E\left(\underset{(p\times p)}{\begin{bmatrix} (Z_1-\mu_{z,1})^2 & (Z_1-\mu_{z,1})(Z_2-\mu_{z,2}) & \cdots & (Z_1-\mu_{z,1})(Z_p-\mu_{z,p}) \\ (Z_2-\mu_{z,2})(Z_1-\mu_{z,1}) & (Z_2-\mu_{z,2})^2 & \cdots & (Z_2-\mu_{z,2})(Z_p-\mu_{z,p}) \\ \vdots & \vdots & \ddots & \vdots \\ (Z_p-\mu_{z,p})(Z_1-\mu_{z,1}) & (Z_p-\mu_{z,p})(Z_2-\mu_{z,2}) & \cdots & (Z_p-\mu_{z,p})^2 \end{bmatrix}}\right)$$

$$= E\underset{(p\times p)}{\begin{bmatrix} (Z_1-\mu_{z,1})^2 & (Z_1-\mu_{z,1})(Z_2-\mu_{z,2}) & \cdots & (Z_1-\mu_{z,1})(Z_p-\mu_{z,p}) \\ (Z_2-\mu_{z,2})(Z_1-\mu_{z,1}) & (Z_2-\mu_{z,2})^2 & \cdots & (Z_2-\mu_{z,2})(Z_p-\mu_{z,p}) \\ \vdots & \vdots & \ddots & \vdots \\ (Z_p-\mu_{z,p})(Z_1-\mu_{z,1}) & (Z_p-\mu_{z,p})(Z_2-\mu_{z,2}) & \cdots & (Z_p-\mu_{z,p})^2 \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} E(Z_1-\mu_{z,1})^2 & E(Z_1-\mu_{z,1})(Z_2-\mu_{z,2}) & \cdots & E(Z_1-\mu_{z,1})(Z_p-\mu_{z,p}) \\ E(Z_2-\mu_{z,2})(Z_1-\mu_{z,1}) & E(Z_2-\mu_{z,2})^2 & \cdots & E(Z_2-\mu_{z,2})(Z_p-\mu_{z,p}) \\ \vdots & \vdots & \ddots & \vdots \\ E(Z_p-\mu_{z,p})(Z_1-\mu_{z,1}) & E(Z_p-\mu_{z,p})(Z_2-\mu_{z,2}) & \cdots & E(Z_p-\mu_{z,p})^2 \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \sigma_{z,11} & \sigma_{z,12} & \cdots & \sigma_{z,1p} \\ \sigma_{z,21} & \sigma_{z,22} & \cdots & \sigma_{z,2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{z,p1} & \sigma_{z,p2} & \cdots & \sigma_{z,pp} \end{bmatrix}} = \underset{(p\times p)}{\begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{11}}} & \frac{\sigma_{p2}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pp} \end{bmatrix}} = \underset{(p\times p)}{\begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}} = \underset{(p\times p)}{\boldsymbol{\rho}}$$

Hence, the population variance-covariance matrix for $\underset{(p\times 1)}{\mathbf{Z}}$ is equal to the population correlation matrix of $\underset{(p\times 1)}{\mathbf{X}}$. That is, $\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ ■

3.5 Mean Vector and Variance-Covariance Matrix for Linear Combinations of Standardized Continuous Random Variables

3.5.1 Linear Combination of Standardized Continuous Random Variables

Definition 3.5.1 (Linear Combination of **Z**). *Let* $\underset{(p\times 1)}{\mathbf{c}}$ *be a* $p \times 1$ *vector of constants defined as*

$$\underset{(p\times 1)}{\mathbf{c}} = \underset{(p\times 1)}{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}}$$

and let $\underset{(p\times 1)}{\mathbf{Z}}$ *be a* $p \times 1$ *population random vector of standardized continuous random variables*

$$\underset{(p\times 1)}{\mathbf{Z}} = \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}}$$

Then a **linear combination of** $\underset{(p\times 1)}{\mathbf{Z}}$, p *standardized random variables, is given by the inner product*

$$\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = \underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}} = c_1Z_1 + c_2Z_2 + \cdots + c_pZ_p.$$

3.5.2 Population Parameters for Linear Combinations of Standardized Continuous Random Variables

Theorem 3.5.1 (Mean of a Linear Combination of **Z**). *Suppose a linear combination of* $\underset{(p\times 1)}{\mathbf{Z}}$, $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$, *is given by Definition 3.5.1 and a population mean vector* of $\underset{(p\times 1)}{\mathbf{Z}}$, $\underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} = \underset{(p\times 1)}{E(\mathbf{Z})} = \underset{(p\times 1)}{\mathbf{0}}$, *is given by Definition 3.4.1. Then the* **expected value** *or* **mean of a linear combination** of $\underset{(p\times 1)}{\mathbf{Z}}$, *is given by*

$$E\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right) = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} = 0$$

Proof. Using linearity of E and Theorem 3.4.1.

$$E\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right)$$

$$= E\left(\underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}}\right)$$

$$= E(c_1Z_1 + c_2Z_2 + \cdots + c_pZ_p)$$

$$= c_1E(Z_1) + c_2E(Z_2) + \cdots + c_pE(Z_p)$$

$$= c_1\mu_{z,1} + c_2\mu_{z,2} + \cdots + c_p\mu_{z,p}$$

$$= \underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \underset{(p\times 1)}{\begin{bmatrix} \mu_{z,1} \\ \mu_{z,2} \\ \vdots \\ \mu_{z,p} \end{bmatrix}}$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}}$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{0}}$$

$$= 0 \ \blacksquare$$

Theorem 3.5.2 (Variance of a Linear Combination of **Z**). *Suppose a linear combination of* $\underset{(p\times 1)}{\mathbf{Z}}$, $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$, *is given by Definition 3.5.1 and a population variance-covariance of* $\underset{(p\times 1)}{\mathbf{Z}}$, $\underset{(p\times p)}{\Sigma_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$, *is given by Theorem 3.4.4. Then the* **variance of a linear combination of** $\underset{(p\times 1)}{\mathbf{Z}}$, *is given by*

$$\mathrm{var}\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right) = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}} = \sum_{i}^{p} \sum_{k=1}^{p} c_i c_k \sigma_{z,ik}$$

$$= \sum_{i=1}^{p} c_i^2 \sigma_{z,ii} + \underbrace{\sum \sum c_i c_k \sigma_{z,ik}}_{i \neq k} = \sum_{i=1}^{p} c_i^2 \sigma_{z,ii} + 2 \sum_{i<k} c_i c_k \sigma_{z,ik}$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}} = \sum_{i}^{p} \sum_{k=1}^{p} c_i c_k \rho_{ik}$$

$$= \sum_{i=1}^{p} c_i^2 \rho_{ii} + \underbrace{\sum \sum c_i c_k \rho_{ik}}_{i \neq k} = p c_i^2 + 2 \sum_{i<k} c_i c_k \rho_{ik}.$$

Proof. Follows directly from Theorem 3.3.2, Theorem 3.4.3, and Theorem 3.4.4 ■

Theorem 3.5.3 (Covariance of Two Linear Combinations of **Z**). *Suppose two linear combinations of* $\underset{(p\times 1)}{\mathbf{Z}}$, $\underset{(1\times p)}{\mathbf{b}'}\cdot\underset{(p\times 1)}{\mathbf{Z}}$ and $\underset{(1\times p)}{\mathbf{c}'}\cdot\underset{(p\times 1)}{\mathbf{Z}}$, *are given following Definition 3.5.1 and a population variance-covariance of* $\underset{(p\times 1)}{\mathbf{Z}}$, $\underset{(p\times p)}{\Sigma_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ *is given by Theorem 3.4.4.Then the* **covariance of two linear combinations of** $\underset{(p\times 1)}{\mathbf{Z}}$, *is given by*

$$
\begin{aligned}
\operatorname{cov}\left(\underset{(1\times p)}{\mathbf{b}'}\cdot\underset{(p\times 1)}{\mathbf{Z}},\underset{(1\times p)}{\mathbf{c}'}\cdot\underset{(p\times 1)}{\mathbf{Z}}\right) &= \underset{(1\times p)}{\mathbf{b}'}\cdot\underset{(p\times p)}{\Sigma_{\mathbf{Z}}}\cdot\underset{(p\times 1)}{\mathbf{c}} = \sum_{i=1}^{p}\sum_{k=1}^{p} b_i c_k \sigma_{z,ik} \\
&= \sum_{i=1}^{p}\sum_{k=1}^{p} b_i c_k \sigma_{z,ik} = p b_i c_i + \underbrace{\sum\sum b_i c_k \sigma_{z,ik}}_{i\neq k} \\
&= p b_i c_i + \underbrace{\sum\sum b_i c_k \rho_{ik}}_{i\neq k} = \sum_{i=1}^{p}\sum_{k=1}^{p} b_i c_k \rho_{ik} \\
&= \underset{(1\times p)}{\mathbf{b}'}\cdot\underset{(p\times p)}{\boldsymbol{\rho}}\cdot\underset{(p\times 1)}{\mathbf{c}}.
\end{aligned}
$$

Proof. Follows directly from Theorem 3.3.3, Theorem 3.4.3, and Theorem 3.4.4 ■

3.5.3 q Linear Combinations of Standardized Continuous Random Variables

Definition 3.5.2 (q Linear Combinations of $\mathbf{Z}$). *Consider* $\underset{(q\times p)}{\mathbf{C}}$ *a matrix of real constants and the* $\boldsymbol{q}$ **linear combinations of** $\underset{(p\times 1)}{\mathbf{Z}}$, Y_i,

$$Y_1 = \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = c_{11}Z_1 + c_{12}Z_2 + \cdots + c_{1p}Z_p$$

$$Y_2 = \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = c_{21}Z_1 + c_{22}Z_2 + \cdots + c_{2p}Z_p$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$Y_q = \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = c_{q1}Z_1 + c_{q2}Z_2 + \cdots + c_{qp}Z_p$$

or in matrix notation,

$$\underset{(q\times 1)}{\mathbf{Y}} = \underset{(q\times 1)}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{bmatrix}} = \underset{(q\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{Z}} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{Z}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\mathbf{Z}} \end{bmatrix}} = \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}} = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{Z}}$$

3.5.4 Population Mean Vector for q Linear Combinations of Standardized Continuous Random Variables

Theorem 3.5.4 (Population Mean Vector for q Linear Combinations of $\mathbf{Z}$). *Suppose* q *linear combinations of* $\underset{(p\times 1)}{\mathbf{Z}}$, $Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$ *are given by Definition 3.5.2 and a population mean vector of* $\underset{(p\times 1)}{\mathbf{Z}}$, $\underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} = \underset{(p\times 1)}{E(\mathbf{Z})} = \underset{(p\times 1)}{\mathbf{0}}$ *is given by Definition 3.4.1. Then the* **population mean vector for q linear combinations of** $\underset{(p\times 1)}{\mathbf{Z}}$, $\underset{(p\times 1)}{\mathbf{Y}}$, *is given by*

$$\underset{(q\times 1)}{\boldsymbol{\mu}_{\mathbf{Y}}} = \underset{(q\times 1)}{E(\mathbf{Y})} = E\left(\underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right) = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} = \underset{(q\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} \end{bmatrix}} = \underset{(q\times 1)}{\mathbf{0}}$$

Proof. Using the linearity of E and Definition 2.2.5.

$\underset{(q\times 1)}{\boldsymbol{\mu}_{\mathbf{Y}}}$

$$= \underset{(q\times 1)}{E(\mathbf{Y})}$$

$$= E\left(\underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right)$$

$$= E\left(\underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}}\right)$$

$$= \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} E\left(\underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}} \right)$$

$$= \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} E(Z_1) \\ E(Z_2) \\ \vdots \\ E(Z_p) \end{bmatrix}}$$

$$= \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} \mu_{z,1} \\ \mu_{z,2} \\ \vdots \\ \mu_{z,p} \end{bmatrix}}$$

$$= \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}}$$

$$= \underset{(q\times 1)}{\begin{bmatrix} c_{11}\mu_{z,1} + c_{12}\mu_{z,2} + \cdots + c_{1p}\mu_{z,p} \\ c_{21}\mu_{z,1} + c_{22}\mu_{z,2} + \cdots + c_{2p}\mu_{z,p} \\ \vdots \\ c_{q1}\mu_{z,1} + c_{q2}\mu_{z,2} + \cdots + c_{qp}\mu_{z,p} \end{bmatrix}}$$

$$= \underset{(q\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}'_1} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} \\ \underset{(1\times p)}{\mathbf{c}'_2} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{c}'_q} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} \end{bmatrix}}$$

$$= \underset{(q\times 1)}{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}$$

$$= \underset{(q\times 1)}{\mathbf{0}} \;\blacksquare$$

Thus, the ith row of $\underset{(p\times 1)}{\mathbf{Y}}$ has population mean

$$\bar{Y}_i = E(Y_i) = E\left(\underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{Z}}} = 0$$

for $i = 1,2, \dots, q$.

3.5.5 Population Variance-Covariance Matrix for q Linear Combinations of Standardized Continuous Random Variables

Theorem 3.5.5. (Population Variance-Covariance Matrix for q Linear Combinations of $\mathbf{Z}$). *Suppose* q *linear combinations of* $\underset{(p\times 1)}{\mathbf{Z}}$, $Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$ *are given by Definition 3.5.2 and a population variance-covariance of* $\underset{(p\times 1)}{\mathbf{Z}}$, $\underset{(p\times p)}{\Sigma_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ *is given by Theorem 3.4.4. Then the symmetric* **population variance-covariance matrix for** q **linear combinations of** $\underset{(p\times 1)}{\mathbf{Z}}$, $\underset{(p\times 1)}{\mathbf{Y}}$, *is given by*

$$\underset{(q\times q)}{\Sigma_{\mathbf{Y}}} = \underset{(q\times q)}{\mathrm{Cov}(\mathbf{Y})} = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{Z}}} \cdot \underset{(p\times q)}{\mathbf{C}'} = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times q)}{\mathbf{C}'}$$

Proof. Using Definition 2.2.5 for matrix multiplication and following Theorem 3.5.2 for computation of diagonal elements and Theorem 3.5.3 for computation of off-diagonal elements.

$$\underset{(q\times q)}{\Sigma_{\mathbf{Y}}}$$

$$= \underset{(q\times q)}{\mathrm{Cov}(\mathbf{Y})}$$

$$= \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{Z}}} \cdot \underset{(p\times q)}{\mathbf{C}'}$$

$$= \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times p)}{\begin{bmatrix} \sigma_{z,11} & \sigma_{z,12} & \cdots & \sigma_{z,1p} \\ \sigma_{z,21} & \sigma_{z,22} & \cdots & \sigma_{z,2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{z,p1} & \sigma_{z,p2} & \cdots & \sigma_{z,pp} \end{bmatrix}} \underset{(p\times q)}{\begin{bmatrix} c_{11} & c_{21} & \cdots & c_{q1} \\ c_{12} & c_{22} & \cdots & c_{q2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1p} & c_{2p} & \cdots & c_{qp} \end{bmatrix}}$$

$$= \begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \end{bmatrix}_{(q\times q)}$$

$$= \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times q)}{\mathbf{C}'}$$

$$= \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times p)}{\begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pp} \end{bmatrix}} \underset{(p\times q)}{\begin{bmatrix} c_{11} & c_{21} & \cdots & c_{q1} \\ c_{12} & c_{22} & \cdots & c_{q2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1p} & c_{2p} & \cdots & c_{qp} \end{bmatrix}}$$

$$= \begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \end{bmatrix}_{(q\times q)} \blacksquare$$

Thus, the ith row $\underset{(q\times 1)}{\mathbf{Y}}$ has population variance

$$\text{var}(Y_i) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_i} = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_i}$$

for $i = 1,2, \dots, q$.

And the ith row and kth row of $\underset{(q\times 1)}{\mathbf{Y}}$ have population covariance

$$\begin{aligned} &\text{cov}(Y_i, Y_k) \\ &= \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_k} = \underset{(1\times p)}{\mathbf{c}_k'} \cdot \underset{(p\times p)}{\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_i} \\ &= \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_k} = \underset{(1\times p)}{\mathbf{c}_k'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{c}_i} \end{aligned}$$

for $i, k = 1,2, \dots, q$.

Chapter 4

Multivariate Sample Theory

4.1 Organization of Multivariate Sample Data

Multivariate sample data arise whenever an investigator, seeking to understand a social or physical phenomenon, selects a number $p > 1$ of **variables** or **characteristics** to record. The values of these variables are all recorded for each distinct **multivariate observation**.

We will use the notation x_{jk}, for realized samples, to indicate the particular value of the kth variable (characteristic) on the jth multivariate observation. That is,

$$x_{jk} = \text{measurement of the } k\text{th variable on the } j\text{th multivariate observation}$$

Consequently, n multivariate observations on p variables (characteristic) can be displayed as follows:

$$\begin{array}{ccccccc}
 & \text{Variable 1} & \text{Variable 2} & \cdots & \text{Variable } k & \cdots & \text{Variable } p \\
\text{Observation 1:} & x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\
\text{Observation 2:} & x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\
\vdots & \vdots & \vdots & & \vdots & & \vdots \\
\text{Observation } j\text{:} & x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\
\vdots & \vdots & \vdots & & \vdots & & \vdots \\
\text{Observation } n\text{:} & x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np}
\end{array}$$

for $j = 1,2, \ldots, n$ multivariate observations and $k = 1,2, \ldots, p$ variables [3, p. 5].

A variable or column of the multivariate sample data array is called a realized **characteristic vector** of dimension $n \times 1$

$$\underset{(n\times 1)}{\mathbf{x}_k} = \underset{(n\times 1)}{\begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{nk} \end{bmatrix}}$$

for $k = 1,2,\dots,p$. Where the transpose of the characteristic vector is of dimension $1 \times n$

$$\underset{(1\times n)}{\mathbf{x}'_k} = \underset{(1\times n)}{[x_{1k}, x_{2k}, \dots, x_{nk}]}.$$

A realized **multivariate observation vector** of dimension $p \times 1$ is given by

$$\underset{(p\times 1)}{\mathbf{x}_j} = \underset{(p\times 1)}{\begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jp} \end{bmatrix}}$$

for $j = 1,2,\dots,n$. Where a row of the multivariate sample data array is given by the transpose of a multivariate observation vector of dimension $1 \times p$

$$\underset{(1\times p)}{\mathbf{x}'_j} = \underset{(1\times p)}{\left[x_{j1}, x_{j2}, \dots, x_{jp}\right]}.$$

The $n \times p$ **multivariate sample matrix** $\underset{(n\times p)}{\mathbf{X}}$ can also be displayed as n realized transposed multivariate observation vectors, stacked on top of each-other, each with p characteristics or variable elements.

$$\underset{(n\times p)}{\mathbf{X}} = \underset{(n\times p)}{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix}} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_j' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix}$$

for $j = 1,2, \dots, n$ multivariate observations and $k = 1,2, \dots, p$ variables.

4.2 Random Samples

4.2.1 Univariate Random Sample

Definition 4.2.1 (Univariate Random Sample). *If random variables X_{jk} for $j = 1,2, \dots, n$ are independent and identically distributed (iid) from a common population continuous random variable X_k, with univariate marginal pdf $f_k(x_k)$, population mean μ_k, and population variance σ_{kk}; then, $X_{1k}, X_{2k}, \dots, X_{nk}$ constitute a* **univariate random sample** *of size n* [6, p. 226].

One should be aware that the elements X_{jk} for $j = 1,2, \dots, n$ must be independent; however, random variables (characteristics) X_k from $k = 1,2, \dots, p$ are generally not assumed independent--especially when realized on the same multivariate observations [3, p. 119].

4.2.2 Multivariate Random Sample

Definition 4.2.2 (Multivariate Random Sample). *If random vectors*

$$\underset{(p\times 1)}{\mathbf{X}_j} = \underset{(p\times 1)}{\begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jp} \end{bmatrix}}$$

for $j = 1,2,\dots,n$ *are independent and identically distributed* (iid) *observed from a common population random vector of continuous random variables*

$$\underset{(p\times 1)}{\mathbf{X}} = \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}},$$

defined in Definition 3.21., with joint pdf

$$f\left(\underset{(p\times 1)}{\mathbf{x}}\right) = f_{12\cdots p}(x_1, x_2, \dots, x_p),$$

defined in definition 3.2.2., population mean vector

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

defined in Definition 3.2.11., and population variance-covariance matrix

$$\underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} = \underset{(p\times p)}{\mathrm{Cov}(\mathbf{X})} = E\underset{(p\times 1)}{(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})}\underset{(1\times p)}{(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})'};$$

defined in Theorem 3.2.1.,then, these random vectors $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \dots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a* **multivariate random sample** *of size* n *from a* $p - variate$ *population.*

4.2.3 Multivariate Random Sample Matrix

Definition 4.2.3 (Multivariate Random Sample Matrix). *A* **multivariate random sample matrix** *is a random matrix whose row vectors are unrealized multivariate sample observations*

$$\underset{(1\times p)}{\mathbf{X}_j'} = \underset{(1\times p)}{\left[X_{j1}, X_{j2}, \dots, X_{jp}\right]}$$

for $j = 1,2,\dots,n$. *In addition, the column vectors of the matrix are unrealized variables or characteristics taken on each of the* n *multivariate sample observations*

$$\underset{(n\times 1)}{\mathbf{X}_k} = \underset{(n\times 1)}{\begin{bmatrix} X_{1k} \\ X_{2k} \\ \vdots \\ X_{nk} \end{bmatrix}}$$

for $k = 1,2,\dots,p$. *Let the* $(j,k)th$ *entry be a continuous random variable* X_{jk}, *then the* $n\times p$ *multivariate random sample matrix* $\underset{(n\times p)}{\mathbf{X}} = \{X_{jk}\}$ *is defined by*

$$\underset{(n\times p)}{\mathbf{X}} = \underset{(n\times p)}{\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2k} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{j1} & X_{j2} & \cdots & X_{jk} & \cdots & X_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} & \cdots & X_{np} \end{bmatrix}} = \begin{bmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \\ \vdots \\ \mathbf{X}_j' \\ \vdots \\ \mathbf{X}_n' \end{bmatrix}$$

for $j = 1,2,\dots,n$ *and* $k = 1,2,\dots,p$. *Since the row vectors* $\underset{(1\times p)}{\mathbf{X}_1'}, \underset{(1\times p)}{\mathbf{X}_2'}, \dots, \underset{(1\times p)}{\mathbf{X}_n'}$ *represent* iid *multivariate sample observations with common joint pdf,* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \dots, \underset{(p\times 1)}{\mathbf{X}_n}$ *are said to form a multivariate random sample* [3, p. 119].

One often refers to each

$$\underset{(p\times 1)}{\mathbf{X}_j} = \underset{(p\times 1)}{\begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jp} \end{bmatrix}}$$

for $j = 1,2, \dots, n$, as an *unrealized* multivariate sample observation (vector). When the multivariate sample observation (vector) has been *realized* (drawn) the notation becomes

$$\underset{(p\times 1)}{\mathbf{x}_j} = \underset{(p\times 1)}{\begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jp} \end{bmatrix}}$$

for $j = 1,2, \dots, n$. Similarly, one often refers to each

$$\underset{(n\times 1)}{\mathbf{X}_k} = \underset{(n\times 1)}{\begin{bmatrix} X_{1k} \\ X_{2k} \\ \vdots \\ X_{nk} \end{bmatrix}}$$

for $k = 1,2, \dots, p$, as an *unrealized* sample characteristic (vector).

When the sample characteristic (vector) has been *realized* the notation becomes

$$\underset{(n\times 1)}{\mathbf{x}_k} = \underset{(n\times 1)}{\begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{nk} \end{bmatrix}}$$

for $k = 1,2, \dots, p$.

4.3 Sample Statistics

Definition 4.3.1 (Sample Mean). *Let* X_{jk} *for* $j = 1,2,\dots,n$ *be iid continuous random variables with common population univariate marginal pdf* $f_k(x_k)$*, mean* μ_k*, and variance* σ_{kk}*. Then the unrealized* **sample mean** $\bar{X}_k$ *is defined by*

$$\bar{X}_k = \frac{1}{n}\sum\nolimits_{j=1}^{n} X_{jk}$$

for $k = 1,2,\dots,p$ *where* $-\infty < \bar{X}_k < \infty$*.*

Because $E(\bar{X}_k) = \mu_k$, one can say $\bar{X}_k$ is an unbiased estimator for the univariate marginal population mean μ_k.

Definition 4.3.2 (Sample Variance). *Let* X_{jk} *for* $j = 1,2,\dots,n$ *be iid continuous random variables with common population univariate marginal pdf* $f_k(x_k)$*, mean* μ_k*, and variance* σ_{kk}*. Then the unrealized* **sample variance** S_{kk} *is defined by*

$$S_{kk} = \frac{1}{n-1}\sum\nolimits_{j=1}^{n}\left(X_{jk} - \bar{X}_k\right)^2$$

for $k = 1,2,\dots,p$ *where* $0 < S_{kk} < \infty$.

Because $E(S_{kk}) = \sigma_{kk}$, one can say S_{kk} is an unbiased estimator for the univariate marginal population variance σ_{kk}. Although, the **sample standard deviation** $\sqrt{S_{kk}}$ is a biased estimator for the univariate marginal population standard deviation $\sqrt{\sigma_{kk}}$; given, $E\left(\sqrt{S_{kk}}\right) \neq \sqrt{\sigma_{kk}}$.

Definition 4.3.3 (Sample Covariance). *Let* $\underset{(2\times1)}{\mathbf{X}_j} = \underset{(2\times1)}{\begin{bmatrix} X_{ji} \\ X_{jk} \end{bmatrix}}$ *for* $j = 1,2,\dots,n$ *be iid continuous random vectors with common population bivariate (joint) marginal pdf* $f_{ik}(x_i, x_k)$*. Denote the common population univariate marginal pdf for* X_{ji} *as* $f_i(x_i)$ *with mean and variance* $\underset{(2\times1)}{\begin{bmatrix} \mu_i \\ \sigma_{ii} \end{bmatrix}}$ *and common population univariate marginal pdf for* X_{jk} *as* $f_k(x_k)$ *with mean and variance* $\underset{(2\times1)}{\begin{bmatrix} \mu_k \\ \sigma_{kk} \end{bmatrix}}$*. Then the unrealized* **sample covariance** S_{ik} *is defined by*

$$S_{ik} = \frac{1}{n-1}\sum_{j=1}^{n}(X_{ji} - \bar{X}_i)(X_{jk} - \bar{X}_k)$$

for $i, k = 1,2,\dots,p$ *where* $-\infty < S_{ik} < \infty$*.*

Because $E(S_{ik}) = \sigma_{ik}$, one can say S_{ik} is an unbiased estimator for the bivariate marginal population covariance σ_{ik}. Given that $\underset{(2\times1)}{\mathbf{X}_j} = \underset{(2\times1)}{\begin{bmatrix} X_{ji} \\ X_{jk} \end{bmatrix}} \subseteq \underset{(p\times1)}{\begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jp} \end{bmatrix}}$ for $j = 1,2,\dots,n$, the multivariate random sample is collected on p characteristics and then subset into bivariate pairs. Furthermore, $S_{ik} = S_{ki}$ and when $i = k$ the sample covariance becomes the sample variance S_{kk}.

Definition 4.3.4 (Sample Correlation). *Let* $\underset{(2\times1)}{\mathbf{X}_j} = \underset{(2\times1)}{\begin{bmatrix} X_{ji} \\ X_{jk} \end{bmatrix}}$ *for* $j = 1,2,\dots,n$ *be iid continuous random vectors with common population bivariate (joint) marginal pdf* $f_{ik}(x_i, x_k)$. *Denote the common population univariate marginal pdf for* X_{ji} *as* $f_i(x_i)$ *with mean and variance* $\underset{(2\times1)}{\begin{bmatrix} \mu_i \\ \sigma_{ii} \end{bmatrix}}$ *and common population univariate marginal pdf for* X_{jk} *as* $f_k(x_k)$ *with mean and variance* $\underset{(2\times1)}{\begin{bmatrix} \mu_k \\ \sigma_{kk} \end{bmatrix}}$. *Then the unrealized* **sample correlation** R_{ik} is defined by

$$R_{ik} = \frac{S_{ik}}{\sqrt{S_{ii}}\sqrt{S_{kk}}}$$

$$= \frac{\frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ji} - \bar{X}_i\right)\left(X_{jk} - \bar{X}_k\right)}{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_{ji} - \bar{X}_i)^2}\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_{jk} - \bar{x}_k)^2}}$$

$$= \frac{\sum_{j=1}^{n}\left(X_{ji} - \bar{X}_i\right)\left(X_{jk} - \bar{X}_k\right)}{\sqrt{\sum_{j=1}^{n}(X_{ji} - \bar{X}_i)^2}\sqrt{\sum_{j=1}^{n}(X_{jk} - \bar{X}_k)^2}}$$

for $i, k = 1,2,\dots,p$ where $-1 \le R_{ik} \le 1$.

Because $E(R_{ik}) \neq \rho_{ik}$, one can say R_{ik} is a biased estimator for the bivariate marginal population correlation ρ_{ik}. Next, $R_{ik} = R_{ki}$ and when $i = k$ the sample correlation becomes $R_{kk} = \frac{S_{kk}}{\sqrt{S_{kk}}\sqrt{S_{kk}}} = \frac{S_{kk}}{S_{kk}} = 1$.

4.4 Sample Mean Vector, Variance-Covariance Matrix, and Correlation Matrix

4.4.1 Sample Mean Vector

Theorem 4.4.1 (Sample Mean Vector for **X**). *Let random vectors* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \dots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition* 4.2.2. *Then the* $p \times 1$ *unrealized* **sample mean vector for** $\underset{(n\times p)}{\mathbf{X}}$ *is defined by*

$$\underset{(p\times 1)}{\overline{\mathbf{X}}} = \frac{1}{n}\sum_{j=1}^{n} \underset{(p\times 1)}{\mathbf{X}_j} = \frac{1}{n} \cdot \underset{(p\times n)}{\mathbf{X}'} \cdot \underset{(n\times 1)}{\mathbf{1}} = \underset{(p\times 1)}{\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}}$$

where $-\infty < \bar{X}_k < \infty$, *for* $k = 1,2,\dots,p$ [3, p. 138].

Proof. Use Definition 2.1.4, Definition 2.2.2, Definition 2.2.3, and Definition 4.3.1.

$$\underset{(p\times 1)}{\overline{\mathbf{X}}}$$

$$= \frac{1}{n}\sum_{j=1}^{n} \underset{(p\times 1)}{\mathbf{X}_j}$$

$$= \frac{1}{n}\left(\underset{(p\times 1)}{\mathbf{X}_1} + \underset{(p\times 1)}{\mathbf{X}_2} + \cdots + \underset{(p\times 1)}{\mathbf{X}_n} \right)$$

$$= \frac{1}{n}\left(\underset{(p\times 1)}{\begin{bmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1p} \end{bmatrix}} + \underset{(p\times 1)}{\begin{bmatrix} X_{21} \\ X_{22} \\ \vdots \\ X_{2p} \end{bmatrix}} + \cdots + \underset{(p\times 1)}{\begin{bmatrix} X_{n1} \\ X_{n2} \\ \vdots \\ X_{np} \end{bmatrix}} \right)$$

$$= \frac{1}{n}\underset{(p\times 1)}{\begin{bmatrix} X_{11} + X_{21} + \cdots + X_{n1} \\ X_{12} + X_{22} + \cdots + X_{n2} \\ \vdots \\ X_{1p} + X_{2p} + \cdots + X_{np} \end{bmatrix}}$$

$$= \frac{1}{n}\underset{(p\times 1)}{\begin{bmatrix} \sum_{j=1}^{n} X_{j1} \\ \sum_{j=1}^{n} X_{j2} \\ \vdots \\ \sum_{j=1}^{n} X_{jp} \end{bmatrix}}$$

$$= \underset{(p\times 1)}{\begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n} X_{j1} \\ \frac{1}{n}\sum_{j=1}^{n} X_{j2} \\ \vdots \\ \frac{1}{n}\sum_{j=1}^{n} X_{jp} \end{bmatrix}}$$

$$= \underset{(p\times 1)}{\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}}$$

In terms of matrix operations $\underset{(p\times 1)}{\bar{\mathbf{X}}}$ can be obtained by

$$\underset{(p\times 1)}{\bar{\mathbf{X}}} = \frac{1}{n}\cdot \underset{(p\times n)}{\mathbf{X}'} \cdot \underset{(n\times 1)}{\mathbf{1}}$$

$$= \frac{1}{n}\underset{(p\times n)}{\begin{bmatrix} X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1p} & X_{2p} & \cdots & X_{np} \end{bmatrix}} \cdot \underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}}$$

$$= \frac{1}{n}\underset{(p\times1)}{\begin{bmatrix} X_{11} + X_{21} + \cdots + X_{n1} \\ X_{12} + X_{22} + \cdots + X_{n2} \\ \vdots \\ X_{1p} + X_{2p} + \cdots + X_{np} \end{bmatrix}}$$

$$= \frac{1}{n}\underset{(p\times1)}{\begin{bmatrix} \sum_{j=1}^{n} X_{j1} \\ \sum_{j=1}^{n} X_{j2} \\ \vdots \\ \sum_{j=1}^{n} X_{jp} \end{bmatrix}}$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n} X_{j1} \\ \frac{1}{n}\sum_{j=1}^{n} X_{j2} \\ \vdots \\ \frac{1}{n}\sum_{j=1}^{n} X_{jp} \end{bmatrix}$$

$$= \underset{(p\times1)}{\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}} \blacksquare$$

Because $E\left(\underset{(p\times1)}{\bar{\mathbf{X}}}\right) = \underset{(p\times1)}{\boldsymbol{\mu}_{\mathbf{X}}}$, one can say $\underset{(p\times1)}{\bar{\mathbf{X}}}$ is an unbiased estimator for the population mean vector $\underset{(p\times1)}{\boldsymbol{\mu}_{\mathbf{X}}}$.

4.4.2 Sample Variance-Covariance Matrix

Theorem 4.4.2 (Sample Variance-Covariance Matrix for **X**). *Let random vectors* $\underset{(p\times1)}{\mathbf{X}_1}, \underset{(p\times1)}{\mathbf{X}_2}, \dots, \underset{(p\times1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition* 4.2.2. *Assume the sample mean vector* $\underset{(p\times1)}{\overline{\mathbf{X}}}$ *defined in Theorem* 4.4.1 *exists. Then the* $p\times p$ *symmetric unrealized* **sample variance-covariance matrix for** $\underset{(n\times p)}{\mathbf{X}}$ *is defined by*

$$\underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}} = \frac{1}{n-1}\sum_{j=1}^{n}\left(\underset{(p\times1)}{\mathbf{X}_j} - \underset{(p\times1)}{\overline{\mathbf{X}}}\right)\left(\underset{(p\times1)}{\mathbf{X}_j} - \underset{(p\times1)}{\overline{\mathbf{X}}}\right)'$$

$$= \frac{1}{n-1}\cdot\left(\underset{(n\times p)}{\mathbf{X}} - \frac{1}{n}\cdot\underset{(n\times1)}{\mathbf{1}}\cdot\underset{(1\times n)}{\mathbf{1}'}\cdot\underset{(n\times p)}{\mathbf{X}}\right)'\cdot\left(\underset{(n\times p)}{\mathbf{X}} - \frac{1}{n}\cdot\underset{(n\times1)}{\mathbf{1}}\cdot\underset{(1\times n)}{\mathbf{1}'}\cdot\underset{(n\times p)}{\mathbf{X}}\right)$$

$$= \frac{1}{n-1}\cdot\left(\underset{(n\times p)}{\mathbf{X}} - \underset{(n\times1)}{\mathbf{1}}\cdot\underset{(1\times p)}{\overline{\mathbf{X}}'}\right)'\cdot\left(\underset{(n\times p)}{\mathbf{X}} - \underset{(n\times1)}{\mathbf{1}}\cdot\underset{(1\times p)}{\overline{\mathbf{X}}'}\right) = \underset{(p\times p)}{\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix}}$$

[3, pp. 123,138].

Proof. Use Definition 2.1.4, Definition 2.2.2, Definition 2.2.3, Definition 2.2.3, Definition 4.2.3, Definition 4.3.2, and Definition 4.3.3.

$$\underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}} = \frac{1}{n-1}\sum_{j=1}^{n}\left(\underset{(p\times1)}{\mathbf{X}_j} - \underset{(p\times1)}{\overline{\mathbf{X}}}\right)\left(\underset{(p\times1)}{\mathbf{X}_j} - \underset{(p\times1)}{\overline{\mathbf{X}}}\right)'$$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(\underset{(p\times1)}{\begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jp} \end{bmatrix}} - \underset{(p\times1)}{\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}}\right)\left(\underset{(p\times1)}{\begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jp} \end{bmatrix}} - \underset{(p\times1)}{\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}}\right)'$$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(\underset{(p\times 1)}{\begin{bmatrix} X_{j1}-\bar{X}_1 \\ X_{j2}-\bar{X}_2 \\ \vdots \\ X_{jp}-\bar{X}_p \end{bmatrix}}\right)\left(\underset{(p\times 1)}{\begin{bmatrix} X_{j1}-\bar{X}_1 \\ X_{j2}-\bar{X}_2 \\ \vdots \\ X_{jp}-\bar{X}_p \end{bmatrix}}\right)'$$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(\underset{(p\times 1)}{\begin{bmatrix} X_{j1}-\bar{X}_1 \\ X_{j2}-\bar{X}_2 \\ \vdots \\ X_{jp}-\bar{X}_p \end{bmatrix}}\right)(\underset{(1\times p)}{[X_{j1}-\bar{X}_1, X_{j2}-\bar{X}_2, \dots, X_{jp}-\bar{X}_p]})$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{\sum_{j=1}^{n}\left(X_{j1}-\bar{X}_1\right)^2}{n-1} & \frac{\sum_{j=1}^{n}(X_{j1}-\bar{X}_1)(X_{j2}-\bar{X}_2)}{n-1} & \cdots & \frac{\sum_{j=1}^{n}(X_{j1}-\bar{X}_1)(X_{jp}-\bar{X}_p)}{n-1} \\ \frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)(X_{j1}-\bar{X}_1)}{n-1} & \frac{\sum_{j=1}^{n}\left(X_{j2}-\bar{X}_2\right)^2}{n-1} & \cdots & \frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)(X_{jp}-\bar{X}_p)}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)(X_{j1}-\bar{X}_1)}{n-1} & \frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)(X_{j2}-\bar{X}_2)}{n-1} & \cdots & \frac{\sum_{j=1}^{n}\left(X_{jp}-\bar{X}_p\right)^2}{n-1} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix}}$$

In terms of matrix operations $\underset{(p\times p)}{\mathbf{S_X}}$ can be obtained by

$$\underset{(p\times p)}{\mathbf{S_X}} = \frac{1}{n-1}\cdot\left(\underset{(n\times p)}{\mathbf{X}} - \frac{1}{n}\cdot\underset{(n\times 1)}{\mathbf{1}}\cdot\underset{(1\times n)}{\mathbf{1}'}\cdot\underset{(n\times p)}{\mathbf{X}}\right)'\cdot\left(\underset{(n\times p)}{\mathbf{X}} - \frac{1}{n}\cdot\underset{(n\times 1)}{\mathbf{1}}\cdot\underset{(1\times n)}{\mathbf{1}'}\cdot\underset{(n\times p)}{\mathbf{X}}\right)$$

where

$$\frac{1}{n}\cdot\underset{(n\times 1)}{\mathbf{1}}\cdot\underset{(1\times n)}{\mathbf{1}'}\cdot\underset{(n\times p)}{\mathbf{X}}$$

$$= \frac{1}{n}\cdot\underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}}\cdot\underset{(1\times n)}{[1_1, 1_2, \dots, 1_n]}\cdot\underset{(n\times p)}{\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}}$$

$$= \underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}} \cdot \frac{1}{n} \cdot \underset{(1\times p)}{\left[\sum\nolimits_{j=1}^{n} X_{j1}, \sum\nolimits_{j=1}^{n} X_{j2}, \dots, \sum\nolimits_{j=1}^{n} X_{jp}\right]}$$

$$= \underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}} \cdot \underset{(1\times p)}{\left[\frac{1}{n}\sum\nolimits_{j=1}^{n} X_{j1}, \frac{1}{n}\sum\nolimits_{j=1}^{n} X_{j2}, \dots, \frac{1}{n}\sum\nolimits_{j=1}^{n} X_{jp}\right]}$$

$$= \underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}} \cdot \underset{(1\times p)}{\left[\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p\right]}$$

$$= \underset{(n\times 1)}{\mathbf{1}} \cdot \underset{(1\times p)}{\bar{\mathbf{X}}'}$$

$$= \underset{(n\times p)}{\begin{bmatrix} \bar{X}_1 & \bar{X}_2 & \cdots & \bar{X}_p \\ \bar{X}_1 & \bar{X}_2 & \cdots & \bar{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ \bar{X}_1 & \bar{X}_2 & \cdots & \bar{X}_p \end{bmatrix}}$$

Thus,

$$\underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}} = \frac{1}{n-1} \cdot \left(\underset{(n\times p)}{\mathbf{X}} - \frac{1}{n} \cdot \underset{(n\times 1)}{\mathbf{1}} \cdot \underset{(1\times n)}{\mathbf{1}'} \cdot \underset{(n\times p)}{\mathbf{X}}\right)' \cdot \left(\underset{(n\times p)}{\mathbf{X}} - \frac{1}{n} \cdot \underset{(n\times 1)}{\mathbf{1}} \cdot \underset{(1\times n)}{\mathbf{1}'} \cdot \underset{(n\times p)}{\mathbf{X}}\right)$$

$$= \frac{1}{n-1} \cdot \left(\underset{(n\times p)}{\mathbf{X}} - \underset{(n\times 1)}{\mathbf{1}} \cdot \underset{(1\times p)}{\bar{\mathbf{X}}'}\right)' \cdot \left(\underset{(n\times p)}{\mathbf{X}} - \underset{(n\times 1)}{\mathbf{1}} \cdot \underset{(1\times p)}{\bar{\mathbf{X}}'}\right)$$

$$= \frac{1}{n-1} \cdot \underset{(n\times p)}{\begin{bmatrix} X_{11}-\bar{X}_1 & X_{12}-\bar{X}_2 & \cdots & X_{1p}-\bar{X}_p \\ X_{21}-\bar{X}_1 & X_{22}-\bar{X}_2 & \cdots & X_{2p}-\bar{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1}-\bar{X}_1 & X_{n2}-\bar{X}_2 & \cdots & X_{np}-\bar{X}_p \end{bmatrix}'} \underset{(n\times p)}{\begin{bmatrix} X_{11}-\bar{X}_1 & X_{12}-\bar{X}_2 & \cdots & X_{1p}-\bar{X}_p \\ X_{21}-\bar{X}_1 & X_{22}-\bar{X}_2 & \cdots & X_{2p}-\bar{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1}-\bar{X}_1 & X_{n2}-\bar{X}_2 & \cdots & X_{np}-\bar{X}_p \end{bmatrix}}$$

$$= \frac{1}{n-1} \cdot \underset{(p\times n)}{\begin{bmatrix} X_{11}-\bar{X}_1 & X_{21}-\bar{X}_1 & \cdots & X_{n1}-\bar{X}_1 \\ X_{12}-\bar{X}_2 & X_{22}-\bar{X}_2 & \cdots & X_{n2}-\bar{X}_2 \\ \vdots & \vdots & \ddots & \vdots \\ X_{1p}-\bar{X}_p & X_{2p}-\bar{X}_p & \cdots & X_{np}-\bar{X}_p \end{bmatrix}} \underset{(n\times p)}{\begin{bmatrix} X_{11}-\bar{X}_1 & X_{12}-\bar{X}_2 & \cdots & X_{1p}-\bar{X}_p \\ X_{21}-\bar{X}_1 & X_{22}-\bar{X}_2 & \cdots & X_{2p}-\bar{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1}-\bar{X}_1 & X_{n2}-\bar{X}_2 & \cdots & X_{np}-\bar{X}_p \end{bmatrix}}$$

$$= \begin{bmatrix} \frac{\sum_{j=1}^{n}(X_{j1}-\bar{X}_1)^2}{n-1} & \frac{\sum_{j=1}^{n}(X_{j1}-\bar{X}_1)(X_{j2}-\bar{X}_2)}{n-1} & \cdots & \frac{\sum_{j=1}^{n}(X_{j1}-\bar{X}_1)(X_{jp}-\bar{X}_p)}{n-1} \\ \frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)(X_{j1}-\bar{X}_1)}{n-1} & \frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)^2}{n-1} & \cdots & \frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)(X_{jp}-\bar{X}_p)}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)(X_{j1}-\bar{X}_1)}{n-1} & \frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)(X_{j2}-\bar{X}_2)}{n-1} & \cdots & \frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)^2}{n-1} \end{bmatrix}_{(p\times p)}$$

$$= \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix}_{(p\times p)} \blacksquare$$

The diagonal elements of the sample variance-covariance matrix are the sample variances

$$S_{kk} = (n-1)^{-1}\sum_{j=1}^{n}\left(X_{jk}-\bar{X}_k\right)^2$$

for $k = 1,2,\dots,p, i = k$ where $S_{ii} = S_{kk}$. The off-diagonal elements of the sample variance-covariance matrix are the sample covariances

$$S_{ik} = (n-1)^{-1}\sum_{j=1}^{n}\left(X_{ji}-\bar{X}_i\right)\left(X_{jk}-\bar{X}_k\right)$$

for $i, k = 1,2,\dots,p, i \neq k$ where $S_{ik} = S_{ki}$. Furthermore,

$$\mathrm{tr}(\mathbf{S_X}) = \sum_{k=1}^{p} S_{kk} = S_{11} + S_{22} + \cdots + S_{pp}$$

(total sample variance). Because, $E\left(\underset{(p\times p)}{\mathbf{S_X}}\right) = \underset{(p\times p)}{\Sigma_{\mathbf{X}}}$, one can say $\underset{(p\times p)}{\mathbf{S_X}}$ is an unbiased estimator for the population variance-covariance matrix $\underset{(p\times p)}{\Sigma_{\mathbf{X}}}$.

4.4.3 Sample Standard Deviation Matrix

Definition 4.4.1 (Sample Standard Deviation Matrix for **X**). *Let random vectors* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \ldots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition* 4.2.2. *Assume the sample standard deviations defined in Definition 4.3.2 exists. Then the* $p \times p$ *diagonal unrealized* **sample standard deviation matrix for** $\underset{(n\times p)}{\mathbf{X}}$ *is defined by*

$$\underset{(p\times p)}{\mathbf{D}^{1/2}} = \underset{(p\times p)}{\begin{bmatrix} \sqrt{S_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{S_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{S_{pp}} \end{bmatrix}}$$

with inverse

$$\underset{(p\times p)}{(\mathbf{D}^{1/2})^{-1}} = \underset{(p\times p)}{\mathbf{D}^{-1/2}} = \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{S_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{S_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{S_{pp}}} \end{bmatrix}}$$

[3, p. 139].

4.4.4 Sample Correlation Matrix

Theorem 4.4.3 (Sample Correlation Matrix for **X**). *Let random vectors* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \dots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition* 4.2.2. *Assume the sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}}$ *defined in Theorem4.4.2 exists, and the inverse sample standard deviation matrix defined in Definition 4.4.1 exists. Then the* $p \times p$ *symmetric unrealized* **sample correlation matrix for** $\underset{(n\times p)}{\mathbf{X}}$ *is defined by*

$$\underset{(p\times p)}{\mathbf{R}}$$

$$= \underset{(p\times p)}{\mathbf{D}^{-1/2}} \cdot \underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}} \cdot \underset{(p\times p)}{\mathbf{D}^{-1/2}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} 1 & R_{12} & \cdots & R_{1p} \\ R_{21} & 1 & \cdots & R_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p1} & R_{p2} & \cdots & 1 \end{bmatrix}}$$

[3, p. 139].

Proof. Use Definition 2.2.5 and Definition 4.3.4.

$$\underset{(p\times p)}{\mathbf{R}} = \underset{(p\times p)}{\mathbf{D}^{-1/2}} \cdot \underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}} \cdot \underset{(p\times p)}{\mathbf{D}^{-1/2}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{S_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{S_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{S_{pp}}} \end{bmatrix}} \underset{(p\times p)}{\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix}} \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{S_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{S_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{S_{pp}}} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{S_{11}}{\sqrt{S_{11}}} & \frac{S_{12}}{\sqrt{S_{11}}} & \cdots & \frac{S_{1p}}{\sqrt{S_{11}}} \\ \frac{S_{21}}{\sqrt{S_{22}}} & \frac{S_{22}}{\sqrt{S_{22}}} & \cdots & \frac{S_{2p}}{\sqrt{S_{22}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{S_{p1}}{\sqrt{S_{pp}}} & \frac{S_{p2}}{\sqrt{S_{pp}}} & \cdots & \frac{S_{pp}}{\sqrt{S_{pp}}} \end{bmatrix}} \cdot \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{S_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{S_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{S_{pp}}} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{S_{11}}{\sqrt{S_{11}}\sqrt{S_{11}}} & \frac{S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}} & \cdots & \frac{S_{1p}}{\sqrt{S_{11}}\sqrt{S_{pp}}} \\ \frac{S_{21}}{\sqrt{S_{22}}\sqrt{S_{11}}} & \frac{S_{22}}{\sqrt{S_{22}}\sqrt{S_{22}}} & \cdots & \frac{S_{2p}}{\sqrt{S_{22}}\sqrt{S_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{S_{p1}}{\sqrt{S_{pp}}\sqrt{S_{11}}} & \frac{S_{p2}}{\sqrt{S_{pp}}\sqrt{S_{22}}} & \cdots & \frac{S_{pp}}{\sqrt{S_{pp}}\sqrt{S_{pp}}} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} 1 & R_{12} & \cdots & R_{1p} \\ R_{21} & 1 & \cdots & R_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p1} & R_{p2} & \cdots & 1 \end{bmatrix}} \blacksquare$$

The diagonal elements of the sample correlation matrix are

$$R_{kk} = \frac{S_{kk}}{\sqrt{S_{kk}}\sqrt{S_{kk}}} = \frac{S_{kk}}{S_{kk}} = 1$$

for $k = 1,2,\dots,p, i = k$ where $R_{ii} = R_{kk}$. The off-diagonal elements of the sample correlation matrix are

$$R_{ik} = \frac{S_{ik}}{\sqrt{S_{ii}}\sqrt{S_{kk}}}$$

for $i, k = 1,2,\dots,p, i \neq k$ where $R_{ik} = R_{ki}$.

Furthermore,

$$\operatorname{tr}(\mathbf{R}) = \sum_{k=1}^{p} R_{kk} = 1 + 1 + \cdots + 1 = p$$

(number of characteristics). Because, $E\left(\underset{(p\times p)}{\mathbf{R}}\right) \neq \underset{(p\times p)}{\boldsymbol{\rho}}$, one can say $\underset{(p\times p)}{\mathbf{R}}$ is a biased estimator for the population correlation matrix $\underset{(p\times p)}{\boldsymbol{\rho}}$. Finally,

$$\underset{(p\times p)}{\mathbf{R}} = \underset{(p\times p)}{\mathbf{D}^{-1/2}} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times p)}{\mathbf{D}^{-1/2}} \Rightarrow \underset{(p\times p)}{\mathbf{S_X}} = \underset{(p\times p)}{\mathbf{D}^{1/2}} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times p)}{\mathbf{D}^{1/2}}$$

[3, p. 140].

4.5 Sample Mean Vector and Variance-Covariance Matrix for Linear Combinations of Continuous Random Variables

4.5.1 Linear Combination

Definition 4.5.1 (Linear Combination of **X**). *Let* $\underset{(p\times 1)}{\mathbf{c}}$ *be a* $p \times 1$ *vector of constants defined as*

$$\underset{(p\times 1)}{\mathbf{c}} = \underset{(p\times 1)}{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}}$$

and let $\underset{(p\times 1)}{\mathbf{X}}$ *be a* $p \times 1$ *population random vector of continuous random variables*

$$\underset{(p\times 1)}{\mathbf{X}} = \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}}$$

Now consider a **linear combination of** $\underset{(n\times p)}{\mathbf{X}}$ *of the form*

$$\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}} = \underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \cdot \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = c_1X_1 + c_2X_2 + \cdots + c_pX_p$$

whose unrealized quantity on the jth *multivariate sample observation is*

$$\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} = \underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \cdot \underset{(p\times 1)}{\begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jp} \end{bmatrix}} = c_1X_{j1} + c_2X_{j2} + \cdots + c_pX_{jp}$$

for $j = 1,2,\dots,n$ [3, p. 140].

4.5.2 Sample Statistics for Linear Combinations

Theorem 4.5.1 (Sample Mean of a Linear Combination of **X**). *Let random vectors* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \dots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition* 4.2.2. *Assume the sample mean vector* $\underset{(p\times 1)}{\overline{\mathbf{X}}}$ *defined in Theorem* 4.4.1 *exists. Next, consider a linear combination of the form* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *with* jth *multivariate sample observation* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j}$ *given in Definition* 4.5.1. *Then, the* *unrealized* **sample mean of a linear combination of** $\underset{(n\times p)}{\mathbf{X}}$ *is defined by*

$$\textit{sample mean of } \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}} = E\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\overline{\mathbf{X}}}.$$

[3, p. 140].

Proof. Use Definition 2.1.6, Definition 2.1.11, and Result 2.2.1. (d).

$$E\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right)$$

$$= \frac{1}{n}\sum_{j=1}^{n} \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j}$$

$$= \frac{1}{n}\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_1} + \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_2} + \cdots + \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_n}\right)$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \left[\frac{1}{n}\left(\underset{(p\times 1)}{\mathbf{X}_1} + \underset{(p\times 1)}{\mathbf{X}_2} + \cdots + \underset{(p\times 1)}{\mathbf{X}_n}\right)\right]$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \left[\frac{1}{n}\left(\underset{(p\times 1)}{\begin{bmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1p} \end{bmatrix}} + \underset{(p\times 1)}{\begin{bmatrix} X_{21} \\ X_{22} \\ \vdots \\ X_{2p} \end{bmatrix}} + \cdots + \underset{(p\times 1)}{\begin{bmatrix} X_{n1} \\ X_{n2} \\ \vdots \\ X_{np} \end{bmatrix}}\right)\right]$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \left[\frac{1}{n} \underset{(p\times 1)}{\begin{bmatrix} X_{11} + X_{21} + \cdots + X_{n1} \\ X_{12} + X_{22} + \cdots + X_{n2} \\ \vdots \\ X_{1p} + X_{2p} + \cdots + X_{np} \end{bmatrix}}\right]$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \left[\frac{1}{n} \begin{bmatrix} \sum_{j=1}^{n} X_{j1} \\ \sum_{j=1}^{n} X_{j2} \\ \vdots \\ \sum_{j=1}^{n} X_{jp} \end{bmatrix}_{(p\times 1)} \right]$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n} X_{j1} \\ \frac{1}{n}\sum_{j=1}^{n} X_{j2} \\ \vdots \\ \frac{1}{n}\sum_{j=1}^{n} X_{jp} \end{bmatrix}}$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}}$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\bar{\mathbf{X}}} \blacksquare$$

Theorem 4.5.2 (Sample Variance of a Linear Combination of **X**). *Let random vectors* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \ldots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition* 4.2.2. *Assume the sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}}$ *defined in Theorem* 4.4.2 *exists. Next consider a linear combination of the form* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *with* jth *multivariate sample observation* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j}$ *given in Definition* 4.5.1. *Then the unrealized* **sample variance of a linear combination of** $\underset{(n\times p)}{\mathbf{X}}$ *is defined by*

$$\textit{sample variance of } \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}} = \operatorname{var}\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}}$$

[3, p. 140].

Proof. Use Definition 2.1.11 and Result 2.2.1. (d).

Since,

$$\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} - \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)^2$$

$$= \left(\underset{(1\times p)}{\mathbf{c}'} \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)\right)^2$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j - \overline{\mathbf{X}}}\right) \cdot \underset{(1\times p)}{\mathbf{c}'} \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j - \overline{\mathbf{X}}}\right)$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j - \overline{\mathbf{X}}}\right) \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j - \overline{\mathbf{X}}}\right)' \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)\left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)' \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$\Rightarrow$

$$\operatorname{var}\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right)$$

$$= \frac{1}{n-1}\sum_{j=1}^{n} \underset{(1\times p)}{\mathbf{c}'} \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)\left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)' \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \left[\frac{1}{n-1}\sum\nolimits_{j=1}^{n} \left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)\left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)'\right] \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$$= \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}} \blacksquare$$

Theorem 4.5.3 (Sample Covariance of Two Linear Combinations of **X**). *Let random vectors* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \ldots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition* 4.2.2. *Assume the sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S_X}}$ *defined in Theorem* 4.4.2 *exists. Next consider two linear combinations of the form* $\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *and* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *with* jth *multivariate sample observations* $\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j}$ *and* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j}$*, respectively, given in Definition* 4.5.1. *Then the unrealized* **sample covariance of two linear combinations of** $\underset{(n\times p)}{\mathbf{X}}$ *is defined by*

$$\textit{sample covariance of } \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{X}} \textit{ and } \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}$$

$$= \operatorname{cov}\left(\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{X}}, \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}}$$

[3, pp. 140-141].

Proof. Use Definition 2.1.11 and Result 2.2.1. (d).

Since,

$$\left(\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} - \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\overline{\mathbf{X}}}\right) \cdot \left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} - \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)$$

$$= \left(\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} - \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\overline{\mathbf{X}}}\right) \cdot \left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} - \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)$$

$$= \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\left(\mathbf{X}_j - \overline{\mathbf{X}}\right)} \cdot \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\left(\mathbf{X}_j - \overline{\mathbf{X}}\right)}$$

$$= \underset{(1\times p)}{\mathbf{b}'} \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j - \overline{\mathbf{X}}}\right) \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j - \overline{\mathbf{X}}}\right)' \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$$= \underset{(1\times p)}{\mathbf{b}'} \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)\left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)' \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$\Rightarrow$

$$\operatorname{cov}\left(\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{X}}, \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right)$$

$$= \frac{1}{n-1}\sum_{j=1}^{n} \underset{(1\times p)}{\mathbf{b}'} \cdot \left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)\left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)' \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$$= \underset{(1\times p)}{\mathbf{b}'} \cdot \left[\frac{1}{n-1}\sum\nolimits_{j=1}^{n} \left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)\left(\underset{(p\times 1)}{\mathbf{X}_j} - \underset{(p\times 1)}{\overline{\mathbf{X}}}\right)'\right] \cdot \underset{(p\times 1)}{\mathbf{c}}$$

$$= \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}} \ \blacksquare$$

4.5.3 q Linear Combinations

Definition 4.5.2 (q Linear Combinations of $\mathbf{X}$). *Let random vectors* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \dots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition* 4.2.2. *Now consider* $\boldsymbol{q}$ **linear combinations of** $\underset{(n\times p)}{\mathbf{X}}$ *of the* p *population continuous random variables* $X_1, X_2, \dots, X_p$ *with form:*

$$Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}} = \underset{(1\times p)}{\left[c_{i1}, c_{i2}, \dots, c_{ip}\right]} \cdot \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = c_{i1}X_1 + c_{i2}X_2 + \cdots + c_{ip}X_p$$

for $i = 1,2,\dots,q$ *linear combinations*

$$Y_1 = \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}} = c_{11}X_1 + c_{12}X_2 + \cdots + c_{1p}X_p$$

$$Y_2 = \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}} = c_{21}X_1 + c_{22}X_2 + \cdots + c_{2p}X_p$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$Y_q = \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\mathbf{X}} = c_{q1}X_1 + c_{q2}X_2 + \cdots + c_{qp}X_p$$

[3, pp. 143-144] *or in matrix notation,*

$$\underset{(q\times 1)}{\mathbf{Y}} = \underset{(q\times 1)}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{bmatrix}} = \underset{(q\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\mathbf{X}} \end{bmatrix}} = \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{X}}$$

where the unrealized quantity on the jth *multivariate sample observation,*

$j = 1,2,\ldots,n$, *on the* ith *linear combination,* $i = 1,2,\ldots,q$, *is*

$$Y_{ji} = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} = \underset{(1\times p)}{[c_{i1}, c_{i2}, \ldots, c_{ip}]} \cdot \underset{(p\times 1)}{\begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jp} \end{bmatrix}} = c_{i1}X_{j1} + c_{i2}X_{j2} + \cdots + c_{ip}X_{jp}.$$

4.5.4 Sample Mean Vector for q Linear Combinations

Definition 4.5.3 (Sample Mean Vector for q Linear Combinations of $\mathbf{X}$). *Let random vectors* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \ldots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition* 4.2.2. *Assume the sample mean vector* $\underset{(p\times 1)}{\overline{\mathbf{X}}}$ *defined in Theorem* 4.4.1 *exists. Next, consider* q *linear combinations of the form* $Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *with* jth *multivariate sample observation* $Y_{ji} = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}_j}$ *given in Definition* 4.5.2. *Then the unrealized* **sample mean vector for q linear combinations of** $\underset{(n\times p)}{\mathbf{X}}$ *is defined by*

$\mathit{sample\ mean\ vector\ of}\ \underset{(q\times 1)}{\mathbf{Y}}$

$$\underset{(q\times 1)}{\overline{\mathbf{Y}}} = E\left(\underset{(q\times 1)}{\mathbf{Y}}\right) = E\left(\underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\overline{\mathbf{X}}}$$

[3, p. 144].

Thus, the ith row of $\underset{(q\times 1)}{\mathbf{Y}}$ has unrealized sample mean

$$\bar{Y}_i = E(Y_i) = E\left(\underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\overline{\mathbf{X}}}$$

for $i = 1,2, \dots, q$.

4.5.5 Sample Variance-Covariance Matrix for q Linear Combinations

Definition 4.5.4 (Sample Variance-Covariance Matrix for q Linear Combinations of **X**). *Let random vectors* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \dots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition* 4.2.2. *Assume the sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}}$ *defined in Theorem* 4.4.2 *exists. Next consider* q *linear combinations of the form* $Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *with* jth *multivariate sample observation* $Y_{ji} = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}_j}$ *given in Definition* 4.5.2. *Then the unrealized* **sample variance-covariance matrix for q linear combinations of** $\underset{(n\times p)}{\mathbf{X}}$ *is defined by*

$$\underset{(q\times q)}{\mathbf{S_Y}} = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times q)}{\mathbf{C}'}$$

$$= \underset{(q\times q)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \end{bmatrix}}$$

[3, p. 144].

Thus, the ith row of $\underset{(q\times 1)}{\mathbf{Y}}$ has unrealized sample variance

$$\mathrm{var}(Y_i) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_i}$$

for $i = 1,2, \dots, q$.

And, the ith row and kth row of $\underset{(p\times 1)}{\mathbf{Y}}$ have unrealized sample covariance

$$\mathrm{cov}(Y_i, Y_k) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_k} = \underset{(1\times p)}{\mathbf{c}_k'} \cdot \underset{(p\times p)}{\mathbf{S_X}} \cdot \underset{(p\times 1)}{\mathbf{c}_i}$$

for $i, k = 1,2, \dots, q$.

4.6 Standardized Random Samples

4.6.1 Standardized Univariate Random Sample

Definition 4.6.1 (Standardized Univariate Random Sample). *Let random variables X_{jk} for $j = 1,2, \dots, n$ constitute a univariate random sample defined in Definition* 4.2.1. *Assume the sample mean $\bar{X}_k$ defined in Definition 4.3.1 and sample variance S_{kk} defined in Definition 4.3.2 exist. Then the* **standardized univariate random sample** *is defined by*

$$Z_{jk} = \frac{X_{jk} - \bar{X}_k}{\sqrt{S_{kk}}}$$

for $j = 1,2, \dots, n$. *Hence, $Z_{1k}, Z_{2k}, \dots, Z_{nk}$ constitute standardized univariate random sample of size n.*

Definition 4.6.2 (Standardized Sample Characteristic Vector). *Let $\underset{(n\times1)}{\mathbf{X}_k}$ be a sample characteristic vector defined in Definition 4.2.3. Assume the sample mean $\bar{X}_k$ defined in Definition 4.3.1 and sample variance S_{kk} defined in Definition 4.3.2 exist. Then the unrealized* **standardized sample characteristic vector** *is defined by*

$$\underset{(n\times1)}{\mathbf{Z}_k} = \frac{\underset{(n\times1)}{\mathbf{X}_k} - \bar{X}_k \cdot \underset{(n\times1)}{\mathbf{1}}}{\sqrt{S_{kk}}} = \underset{(n\times1)}{\begin{bmatrix} \frac{X_{1k} - \bar{X}_k}{\sqrt{S_{kk}}} \\ \frac{X_{2k} - \bar{X}_k}{\sqrt{S_{kk}}} \\ \vdots \\ \frac{X_{nk} - \bar{X}_k}{\sqrt{S_{kk}}} \end{bmatrix}} = \underset{(n\times1)}{\begin{bmatrix} Z_{1k} \\ Z_{2k} \\ \vdots \\ Z_{nk} \end{bmatrix}}$$

for $k = 1,2, \dots, p$ [3, p. 135].

4.6.2 Standardized Multivariate Random Sample

Theorem 4.6.1 (Standardized Multivariate Random Sample)**.** *Let random vectors*

$$\underset{(p\times1)}{\mathbf{X}_j} = \underset{(p\times1)}{\begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jp} \end{bmatrix}}$$

for $j = 1,2,\dots,n$ *constitute a multivariate random sample defined in Definition* 4.2.2.

Assume the sample mean vector $\underset{(p\times1)}{\bar{\mathbf{X}}}$ defined in *Theorem 4.4.1 and inverse sample standard deviation matrix* $\underset{(p\times p)}{\mathbf{D}^{-1/2}}$ *defined in Definition 4.4.1 exist. Then the* **standardized multivariate random sample** *is defined by*

$$\underset{(p\times1)}{\mathbf{Z}_j} = \underset{(p\times p)}{\mathbf{D}^{-1/2}} \cdot \left(\underset{(p\times1)}{\mathbf{X}_j} - \underset{(p\times1)}{\bar{\mathbf{X}}} \right) = \underset{(p\times1)}{\begin{bmatrix} \dfrac{X_{j1} - \bar{X}_1}{\sqrt{S_{11}}} \\ \dfrac{X_{j2} - \bar{X}_2}{\sqrt{S_{22}}} \\ \vdots \\ \dfrac{X_{jp} - \bar{X}_p}{\sqrt{S_{pp}}} \end{bmatrix}} = \underset{(p\times1)}{\begin{bmatrix} Z_{j1} \\ Z_{j2} \\ \vdots \\ Z_{jp} \end{bmatrix}}$$

for $j = 1,2,\dots,n$. *Hence, random vectors* $\underset{(p\times1)}{\mathbf{Z}_1}, \underset{(p\times1)}{\mathbf{Z}_2}, \dots, \underset{(p\times1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample of size* n [3, p. 449].

Proof. Use Definition 2.2.5.

$$\underset{(p\times1)}{\mathbf{Z}_j} = \underset{(p\times p)}{\mathbf{D}^{-1/2}} \cdot \left(\underset{(p\times1)}{\mathbf{X}_j} - \underset{(p\times1)}{\bar{\mathbf{X}}} \right)$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{S_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{S_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{S_{pp}}} \end{bmatrix}} \cdot \left(\underset{(p\times 1)}{\begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jp} \end{bmatrix}} - \underset{(p\times 1)}{\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}} \right)$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{1}{\sqrt{S_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{S_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{S_{pp}}} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} X_{j1} - \bar{X}_1 \\ X_{j2} - \bar{X}_2 \\ \vdots \\ X_{jp} - \bar{X}_p \end{bmatrix}}$$

$$= \underset{(p\times 1)}{\begin{bmatrix} \frac{X_{j1} - \bar{X}_1}{\sqrt{S_{11}}} \\ \frac{X_{j2} - \bar{X}_2}{\sqrt{S_{22}}} \\ \vdots \\ \frac{X_{jp} - \bar{X}_p}{\sqrt{S_{pp}}} \end{bmatrix}}$$

$$= \underset{(p\times 1)}{\begin{bmatrix} Z_{j1} \\ Z_{j2} \\ \vdots \\ Z_{jp} \end{bmatrix}} \text{ for } j = 1,2,\dots,n \blacksquare$$

4.6.3 Standardized Multivariate Random Sample Matrix

Definition 4.6.3 (Standardized Multivariate Random Sample Matrix). *A* **standardized multivariate random sample matrix** *is a matrix whose row vectors are transposed unrealized standardized multivariate random sample observations*

$$\underset{(1\times p)}{\mathbf{Z}_j'} = \underset{(1\times p)}{\left[Z_{j1}, Z_{j2}, \dots, Z_{jp}\right]}$$

for $j = 1,2,\dots,n$ *defined in Theorem* 4.6.1*. In addition, the column vectors of the matrix are unrealized standardized sample variables or characteristic vectors*

$$\underset{(n\times 1)}{\mathbf{Z}_k} = \underset{(n\times 1)}{\begin{bmatrix} Z_{1k} \\ Z_{2k} \\ \vdots \\ Z_{nk} \end{bmatrix}}$$

for $k = 1,2,\dots,p$ *defined in Definition* 4.6.2. *Let the* $(j,k)th$ *entry be a standardized continuous random variable* Z_{jk}*, then the* $n\times p$ *standardized multivariate random sample matrix* $\underset{(n\times p)}{\mathbf{Z}} = \{Z_{jk}\}$ *is defined by*

$$\underset{(n\times p)}{\mathbf{Z}} = \underset{(n\times p)}{\begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1k} & \cdots & Z_{1p} \\ Z_{21} & Z_{22} & \cdots & Z_{2k} & \cdots & Z_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ Z_{j1} & Z_{j2} & \cdots & Z_{jk} & \cdots & Z_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nk} & \cdots & Z_{np} \end{bmatrix}}$$

$$= \underset{(n\times p)}{\begin{bmatrix} \frac{X_{11}-\overline{X}_1}{\sqrt{S_{11}}} & \frac{X_{12}-\overline{X}_2}{\sqrt{S_{22}}} & \cdots & \frac{X_{1k}-\overline{X}_k}{\sqrt{S_{kk}}} & \cdots & \frac{X_{1p}-\overline{X}_p}{\sqrt{S_{pp}}} \\ \frac{X_{21}-\overline{X}_1}{\sqrt{S_{11}}} & \frac{X_{22}-\overline{X}_2}{\sqrt{S_{22}}} & \cdots & \frac{X_{2k}-\overline{X}_k}{\sqrt{S_{kk}}} & \cdots & \frac{X_{2p}-\overline{X}_p}{\sqrt{S_{pp}}} \\ \vdots & \vdots & & \vdots & & \vdots \\ \frac{X_{j1}-\overline{X}_1}{\sqrt{S_{11}}} & \frac{X_{j2}-\overline{X}_2}{\sqrt{S_{22}}} & \cdots & \frac{X_{jk}-\overline{X}_k}{\sqrt{S_{kk}}} & \cdots & \frac{X_{jp}-\overline{X}_p}{\sqrt{S_{pp}}} \\ \vdots & \vdots & & \vdots & & \vdots \\ \frac{X_{n1}-\overline{X}_1}{\sqrt{S_{11}}} & \frac{X_{n2}-\overline{X}_2}{\sqrt{S_{22}}} & \cdots & \frac{X_{nk}-\overline{X}_k}{\sqrt{S_{kk}}} & \cdots & \frac{X_{np}-\overline{X}_p}{\sqrt{S_{pp}}} \end{bmatrix}} = \begin{bmatrix} \mathbf{Z}_1' \\ \mathbf{Z}_2' \\ \vdots \\ \mathbf{Z}_j' \\ \vdots \\ \mathbf{Z}_n' \end{bmatrix}$$

for $j = 1,2, \ldots, n$ *standardized multivariate sample observations and* $k = 1,2, \ldots, p$ *standardized sample characteristics* [3, p. 450].

4.7 Sample Statistics for Standardized Samples

Theorem 4.7.1 (Sample Mean for Z_k). *Let* Z_{jk} *for* $j = 1,2, \ldots, n$ *constitute a standardized univariate random sample defined in Definition* 4.6.1. Then the unrealized **sample mean for** Z_k, $\bar{Z}_k$, *is defined by*

$$\bar{Z}_k = \frac{1}{n}\sum_{j=1}^{n} Z_{jk} = \frac{1}{n}\sum_{j=1}^{n} \frac{X_{jk}-\overline{X}_k}{\sqrt{S_{kk}}} = 0$$

for $k = 1,2, \ldots, p$.

Proof.

$\bar{Z}_k$

$$= \frac{1}{n}\sum_{j=1}^{n} Z_{jk}$$

$$= \frac{1}{n}\sum_{j=1}^{n} \frac{X_{jk} - \overline{X}_k}{\sqrt{S_{kk}}}$$

$$= \frac{1}{\sqrt{S_{kk}}}\left[\frac{1}{n}\sum_{j=1}^{n}(X_{jk} - \overline{X}_k)\right]$$

$$= \frac{1}{\sqrt{S_{kk}}}\left[\frac{1}{n}\sum_{j=1}^{n} X_{jk} - \frac{1}{n}\sum_{j=1}^{n} \overline{X}_k\right]$$

$$= \frac{1}{\sqrt{S_{kk}}}\left[\frac{1}{n}\cdot(n \cdot \overline{X}_k) - \frac{1}{n}(n \cdot \overline{X}_k)\right]$$

$$= \frac{1}{\sqrt{S_{kk}}}[\overline{X}_k - \overline{X}_k] = 0 \blacksquare$$

Theorem 4.7.2 (Sample Variance for Z_k). *Let* Z_{jk} *for* $j = 1,2,\dots,n$ *constitute a standardized univariate random sample defined in Definition* 4.6.1*. Then the unrealized* **sample variance for** Z_k, $S_{z,kk}$, *is defined by*

$$S_{z,kk} = \frac{1}{n-1}\sum_{j=1}^{n}\left(Z_{jk} - \bar{Z}_k\right)^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(\frac{X_{jk} - \overline{X}_k}{\sqrt{S_{kk}}}\right)^2 = 1$$

for $k = 1,2,\dots,p$.

$S_{z,kk}$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(Z_{jk} - \bar{Z}_k\right)^2$$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(\frac{X_{jk} - \overline{X}_k}{\sqrt{S_{kk}}} - 0\right)^2$$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(\frac{X_{jk} - \overline{X}_k}{\sqrt{S_{kk}}}\right)^2$$

$$= \frac{1}{S_{kk}} \cdot \left[\frac{1}{n-1}\sum_{j=1}^{n}\left(X_{jk} - \overline{X}_k\right)^2\right]$$

$$= \frac{1}{S_{kk}} \cdot S_{kk} = 1 \blacksquare$$

Theorem 4.7.3 (Sample Covariance for Z_i and Z_k). *Let* $\underset{(2\times1)}{\mathbf{Z}_j} = \underset{(2\times1)}{\begin{bmatrix} X_{ji} \\ X_{jk} \end{bmatrix}}$ *for* $j = 1,2, \dots, n$ *constitute a two-dimensional characteristic subset of the standardized multivariance random sample defined in Theorem* 4.6.1*. Assume the sample means* $\bar{Z}_i, \bar{Z}_k$ *defined in Theorem 4.7.1 and sample variances* $S_{z,ii}, S_{z,kk}$ *defined in Theorem 4.7.2 exist. Then the unrealized* **sample covariance for** Z_i **and** Z_k, $S_{z,ik}$, *is defined by*

$$S_{z,ik} = \frac{1}{n-1}\sum_{j=1}^{n}(Z_{ji} - \bar{Z}_i)(Z_{jk} - \bar{Z}_k) = \frac{1}{n-1}\sum_{j=1}^{n}\left(\frac{X_{ji} - \overline{X}_i}{\sqrt{S_{ii}}}\right)\left(\frac{X_{jk} - \overline{X}_k}{\sqrt{S_{kk}}}\right) = R_{ik}$$

for $i, k = 1,2, \dots, p$.

Proof.

$S_{z,ik}$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(Z_{ji} - \bar{Z}_i\right)\left(Z_{jk} - \bar{Z}_k\right)$$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(\frac{X_{ji} - \overline{X}_i}{\sqrt{S_{ii}}} - 0\right)\left(\frac{X_{jk} - \overline{X}_k}{\sqrt{S_{kk}}} - 0\right)$$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(\frac{X_{ji} - \overline{X}_i}{\sqrt{S_{ii}}}\right)\left(\frac{X_{jk} - \overline{X}_k}{\sqrt{S_{kk}}}\right)$$

$$= \frac{1}{\sqrt{S_{ii}}\sqrt{S_{kk}}} \cdot \left[\frac{1}{n-1}\sum\nolimits_{j=1}^{n}\left(X_{ji}-\overline{X}_i\right)\left(X_{jk}-\overline{X}_k\right)\right]$$

$$= \frac{1}{\sqrt{S_{ii}}\sqrt{S_{kk}}} \cdot S_{ik} = \frac{S_{ik}}{\sqrt{S_{ii}}\sqrt{S_{kk}}} = R_{ik} \blacksquare$$

Note $S_{z,ik} = S_{z,ki}$, and when $i = k$, $S_{z,kk} = R_{kk} = 1$.

4.8 Sample Mean Vector and Variance-Covariance Matrix for Standardized Samples

4.8.1 Sample Mean Vector for Standardized Samples

Theorem 4.8.1 (Sample Mean Vector for **Z**). *Let* $\underset{(p\times 1)}{\mathbf{Z}_1}, \underset{(p\times 1)}{\mathbf{Z}_2}, \ldots, \underset{(p\times 1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample defined in Theorem* 4.6.1. *Then the* $p \times 1$ *unrealized* **sample mean vector for** $\underset{(n\times p)}{\mathbf{Z}}$ *is defined by*

$$\underset{(p\times 1)}{\bar{\mathbf{Z}}} = \frac{1}{n}\sum\nolimits_{j=1}^{n} \underset{(p\times 1)}{\mathbf{Z}_j} = \frac{1}{n} \cdot \underset{(p\times n)}{\mathbf{Z}'} \cdot \underset{(n\times 1)}{\mathbf{1}} = \underset{(p\times 1)}{\begin{bmatrix} \bar{Z}_1 \\ \bar{Z}_2 \\ \vdots \\ \bar{Z}_p \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}} = \underset{(p\times 1)}{\mathbf{0}}$$

[3, p. 450].

Proof. Use Definition 2.1.4, Definition 2.2.2, Definition 2.2.3, and Theorem 4.7.1.

$$\underset{(p\times 1)}{\bar{\mathbf{Z}}}$$

$$= \frac{1}{n}\sum_{j=1}^{n} \underset{(p\times 1)}{\mathbf{Z}_j}$$

$$= \frac{1}{n}\left(\underset{(p\times 1)}{\mathbf{Z}_1} + \underset{(p\times 1)}{\mathbf{Z}_2} + \cdots + \underset{(p\times 1)}{\mathbf{Z}_n} \right)$$

$$= \frac{1}{n}\left(\underset{(p\times 1)}{\begin{bmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1p} \end{bmatrix}} + \underset{(p\times 1)}{\begin{bmatrix} Z_{21} \\ Z_{22} \\ \vdots \\ Z_{2p} \end{bmatrix}} + \cdots + \underset{(p\times 1)}{\begin{bmatrix} Z_{n1} \\ Z_{n2} \\ \vdots \\ Z_{np} \end{bmatrix}} \right)$$

$$= \frac{1}{n} \underset{(p\times 1)}{\begin{bmatrix} Z_{11} + Z_{21} + \cdots + Z_{n1} \\ Z_{12} + Z_{22} + \cdots + Z_{n2} \\ \vdots \\ Z_{1p} + Z_{2p} + \cdots + Z_{np} \end{bmatrix}}$$

$$= \frac{1}{n} \underset{(p\times 1)}{\begin{bmatrix} \sum_{j=1}^{n} Z_{j1} \\ \sum_{j=1}^{n} Z_{j2} \\ \vdots \\ \sum_{j=1}^{n} Z_{jp} \end{bmatrix}}$$

$$= \underset{(p\times 1)}{\begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n} Z_{j1} \\ \frac{1}{n}\sum_{j=1}^{n} Z_{j2} \\ \vdots \\ \frac{1}{n}\sum_{j=1}^{n} Z_{jp} \end{bmatrix}}$$

$$= \underset{(p\times 1)}{\begin{bmatrix} \bar{Z}_1 \\ \bar{Z}_2 \\ \vdots \\ \bar{Z}_p \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}} = \underset{(p\times 1)}{\mathbf{0}}$$

In terms of matrix operations $\underset{(p\times 1)}{\bar{\mathbf{Z}}}$ can be obtained by

$$\underset{(p\times 1)}{\bar{\mathbf{Z}}} = \frac{1}{n} \cdot \underset{(p\times n)}{\mathbf{Z}'} \cdot \underset{(n\times 1)}{\mathbf{1}}$$

$$= \frac{1}{n} \underset{(p\times n)}{\begin{bmatrix} Z_{11} & Z_{21} & \cdots & Z_{n1} \\ Z_{12} & Z_{22} & \cdots & Z_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{1p} & Z_{2p} & \cdots & Z_{np} \end{bmatrix}} \cdot \underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}}$$

$$= \frac{1}{n} \underset{(p\times 1)}{\begin{bmatrix} Z_{11} + Z_{21} + \cdots + Z_{n1} \\ Z_{12} + Z_{22} + \cdots + Z_{n2} \\ \vdots \\ Z_{1p} + Z_{2p} + \cdots + Z_{np} \end{bmatrix}}$$

$$= \frac{1}{n} \underset{(p\times 1)}{\begin{bmatrix} \sum_{j=1}^{n} Z_{j1} \\ \sum_{j=1}^{n} Z_{j2} \\ \vdots \\ \sum_{j=1}^{n} Z_{jp} \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n} Z_{j1} \\ \frac{1}{n}\sum_{j=1}^{n} Z_{j2} \\ \vdots \\ \frac{1}{n}\sum_{j=1}^{n} Z_{jp} \end{bmatrix}}$$

$$= \underset{(p\times 1)}{\begin{bmatrix} \bar{Z}_1 \\ \bar{Z}_2 \\ \vdots \\ \bar{Z}_p \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}} = \underset{(p\times 1)}{\mathbf{0}} \blacksquare$$

4.8.2 Sample Variance-Covariance Matrix for Standardized Samples

Theorem 4.8.2 (Sample Variance-Covariance Matrix for $\mathbf{Z}$). *Let* $\underset{(p\times 1)}{\mathbf{Z}_1}, \underset{(p\times 1)}{\mathbf{Z}_2}, \dots, \underset{(p\times 1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample defined in Theorem* 4.6.1. *Assume the sample mean vector* $\underset{(p\times 1)}{\bar{\mathbf{Z}}}$ *defined in Theorem 4.8.1 exists.*

Then the $p \times p$ symmetric unrealized **sample variance-covariance matrix for** $\underset{(n\times p)}{\mathbf{Z}}$ *is defined by*

$$\underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} = \frac{1}{n-1}\sum_{j=1}^{n}\left(\underset{(p\times 1)}{\mathbf{Z}_j} - \underset{(p\times 1)}{\bar{\mathbf{Z}}}\right)\left(\underset{(p\times 1)}{\mathbf{Z}_j} - \underset{(p\times 1)}{\bar{\mathbf{Z}}}\right)'$$

$$= \frac{1}{n-1}\cdot\left(\underset{(n\times p)}{\mathbf{Z}} - \frac{1}{n}\cdot\underset{(n\times 1)}{\mathbf{1}}\cdot\underset{(1\times n)}{\mathbf{1}'}\cdot\underset{(n\times p)}{\mathbf{Z}}\right)'\cdot\left(\underset{(n\times p)}{\mathbf{Z}} - \frac{1}{n}\cdot\underset{(n\times 1)}{\mathbf{1}}\cdot\underset{(1\times n)}{\mathbf{1}'}\cdot\underset{(n\times p)}{\mathbf{Z}}\right)$$

$$= \frac{1}{n-1}\cdot\left(\underset{(n\times p)}{\mathbf{Z}} - \underset{(n\times 1)}{\mathbf{1}}\cdot\underset{(1\times p)}{\bar{\mathbf{Z}}'}\right)'\cdot\left(\underset{(n\times p)}{\mathbf{Z}} - \underset{(n\times 1)}{\mathbf{1}}\cdot\underset{(1\times p)}{\bar{\mathbf{Z}}'}\right) = \frac{1}{n-1}\cdot\underset{(p\times n)}{\mathbf{Z}'}\cdot\underset{(n\times p)}{\mathbf{Z}} = \underset{(p\times p)}{\mathbf{R}}$$

[3, p. 450].

Proof. Use Definition 2.1.4, Definition 2.2.2, Definition 2.2.3, Theorem 4.6.1, Theorem 4.7.2, and Theorem 4.7.3.

$$\underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} = \frac{1}{n-1}\sum_{j=1}^{n}\left(\underset{(p\times 1)}{\mathbf{Z}_j} - \underset{(p\times 1)}{\bar{\mathbf{Z}}}\right)\left(\underset{(p\times 1)}{\mathbf{Z}_j} - \underset{(p\times 1)}{\bar{\mathbf{Z}}}\right)'$$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(\underset{(p\times 1)}{\mathbf{Z}_j} - \underset{(p\times 1)}{\mathbf{0}}\right)\left(\underset{(p\times 1)}{\mathbf{Z}_j} - \underset{(p\times 1)}{\mathbf{0}}\right)'$$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\underset{(p\times 1)}{\mathbf{Z}_j}\cdot\underset{(1\times p)}{\mathbf{Z}_j'} = \frac{1}{n-1}\sum_{j=1}^{n}\underset{(p\times 1)}{\begin{bmatrix} Z_{j1} \\ Z_{j2} \\ \vdots \\ Z_{jp}\end{bmatrix}}\underset{(1\times p)}{[Z_{j1}, Z_{j2}, \dots, Z_{jp}]}$$

$$= \frac{1}{n-1}\sum_{j=1}^{n}\left(\underset{(p\times 1)}{\begin{bmatrix} \frac{X_{j1}-\bar{X}_1}{\sqrt{S_{11}}} \\ \frac{X_{j2}-\bar{X}_2}{\sqrt{S_{22}}} \\ \vdots \\ \frac{X_{jp}-\bar{X}_p}{\sqrt{S_{pp}}} \end{bmatrix}} \underset{(1\times p)}{\left[\frac{X_{j1}-\bar{X}_1}{\sqrt{S_{11}}}, \frac{X_{j2}-\bar{X}_2}{\sqrt{S_{22}}}, \dots, \frac{X_{jp}-\bar{X}_p}{\sqrt{S_{pp}}}\right]}\right)$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{\frac{\sum_{j=1}^{n}\left(X_{j1}-\bar{X}_1\right)^2}{\sqrt{S_{11}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{\sum_{j=1}^{n}(X_{j1}-\bar{X}_1)(X_{j2}-\bar{X}_2)}{\sqrt{S_{11}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{\sum_{j=1}^{n}(X_{j1}-\bar{X}_1)(X_{jp}-\bar{X}_p)}{\sqrt{S_{11}}\sqrt{S_{pp}}}}{n-1} \\ \frac{\frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)(X_{j1}-\bar{X}_1)}{\sqrt{S_{22}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)^2}{\sqrt{S_{22}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)(X_{jp}-\bar{X}_p)}{\sqrt{S_{22}}\sqrt{S_{pp}}}}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)(X_{j1}-\bar{X}_1)}{\sqrt{S_{pp}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)(X_{j2}-\bar{X}_2)}{\sqrt{S_{pp}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)^2}{\sqrt{S_{pp}}\sqrt{S_{pp}}}}{n-1} \end{bmatrix}}$$

$$= \underset{(p\times p)}{\begin{bmatrix} \frac{\frac{(n-1)S_{11}}{\sqrt{S_{11}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{(n-1)S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{(n-1)S_{1p}}{\sqrt{S_{11}}\sqrt{S_{pp}}}}{n-1} \\ \frac{\frac{(n-1)S_{21}}{\sqrt{S_{22}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{(n-1)S_{22}}{\sqrt{S_{22}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{(n-1)S_{2p}}{\sqrt{S_{22}}\sqrt{S_{pp}}}}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\frac{(n-1)S_{p1}}{\sqrt{S_{pp}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{(n-1)S_{p2}}{\sqrt{S_{pp}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{(n-1)S_{pp}}{\sqrt{S_{pp}}\sqrt{S_{pp}}}}{n-1} \end{bmatrix}}$$

$$\{S_{ik} = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ji}-\bar{X}_i\right)\left(X_{jk}-\bar{X}_k\right) \Rightarrow (n-1)S_{ik} = \sum_{j=1}^{n}\left(X_{ji}-\bar{X}_i\right)\left(X_{jk}-\bar{X}_k\right) \text{ for } i,k = 1,2,\dots,p\}$$

$$= \begin{bmatrix} \frac{S_{11}}{\sqrt{S_{11}}\sqrt{S_{11}}} & \frac{S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}} & \cdots & \frac{S_{1p}}{\sqrt{S_{11}}\sqrt{S_{pp}}} \\ \frac{S_{21}}{\sqrt{S_{22}}\sqrt{S_{11}}} & \frac{S_{22}}{\sqrt{S_{22}}\sqrt{S_{22}}} & \cdots & \frac{S_{2p}}{\sqrt{S_{22}}\sqrt{S_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{S_{p1}}{\sqrt{S_{pp}}\sqrt{S_{11}}} & \frac{S_{p2}}{\sqrt{S_{pp}}\sqrt{S_{22}}} & \cdots & \frac{S_{pp}}{\sqrt{S_{pp}}\sqrt{S_{pp}}} \end{bmatrix}_{(p\times p)}$$

$$= \begin{bmatrix} 1 & R_{12} & \cdots & R_{1p} \\ R_{21} & 1 & \cdots & R_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p1} & R_{p2} & \cdots & 1 \end{bmatrix}$$

$$= \underset{(p\times p)}{\mathbf{R}}$$

In terms of matrix operations $\underset{(p\times p)}{\mathbf{S_Z}} = \underset{(p\times p)}{\mathbf{R}}$ can be obtained by

$$\underset{(p\times p)}{\mathbf{S_Z}} = \frac{1}{n-1}\cdot\left(\underset{(n\times p)}{\mathbf{Z}} - \frac{1}{n}\cdot\underset{(n\times 1)}{\mathbf{1}}\cdot\underset{(1\times n)}{\mathbf{1}'}\cdot\underset{(n\times p)}{\mathbf{Z}}\right)'\cdot\left(\underset{(n\times p)}{\mathbf{Z}} - \frac{1}{n}\cdot\underset{(n\times 1)}{\mathbf{1}}\cdot\underset{(1\times n)}{\mathbf{1}'}\cdot\underset{(n\times p)}{\mathbf{Z}}\right)$$

where

$$\frac{1}{n}\cdot\underset{(n\times 1)}{\mathbf{1}}\cdot\underset{(1\times n)}{\mathbf{1}'}\cdot\underset{(n\times p)}{\mathbf{Z}}$$

$$= \frac{1}{n}\cdot\underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}}\cdot\underset{(1\times n)}{[1_1, 1_2, \ldots, 1_n]}\cdot\underset{(n\times p)}{\begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1p} \\ Z_{21} & Z_{22} & \cdots & Z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{np} \end{bmatrix}}$$

$$= \underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}}\cdot\frac{1}{n}\cdot\underset{(1\times p)}{\left[\sum_{j=1}^{n} Z_{j1}, \sum_{j=1}^{n} Z_{j2}, \ldots, \sum_{j=1}^{n} Z_{jp}\right]}$$

$$= \underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}}\cdot\frac{1}{n}\cdot\underset{(1\times p)}{\left[\sum_{j=1}^{n}\frac{X_{j1}-\bar{X}_1}{\sqrt{S_{11}}}, \sum_{j=1}^{n}\frac{X_{j2}-\bar{X}_2}{\sqrt{S_{22}}}, \ldots, \sum_{j=1}^{n}\frac{X_{jp}-\bar{X}_p}{\sqrt{S_{pp}}}\right]}$$

$$= \underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}} \cdot \underset{(1\times p)}{\left[\frac{1}{n}\sum_{j=1}^{n}\frac{X_{j1}-\bar{X}_1}{\sqrt{S_{11}}}, \frac{1}{n}\sum_{j=1}^{n}\frac{X_{j2}-\bar{X}_2}{\sqrt{S_{22}}}, \dots, \frac{1}{n}\sum_{j=1}^{n}\frac{X_{jp}-\bar{X}_p}{\sqrt{S_{pp}}}\right]}$$

$$= \underset{(n\times 1)}{\begin{bmatrix} 1_1 \\ 1_2 \\ \vdots \\ 1_n \end{bmatrix}} \cdot \underset{(1\times p)}{\left[\bar{Z}_1, \bar{Z}_2, \dots, \bar{Z}_p\right]}$$

$$= \underset{(n\times 1)}{\mathbf{1}} \cdot \underset{(1\times p)}{\bar{\mathbf{Z}}'}$$

$$= \underset{(n\times p)}{\begin{bmatrix} \bar{Z}_1 & \bar{Z}_2 & \cdots & \bar{Z}_p \\ \bar{Z}_1 & \bar{Z}_2 & \cdots & \bar{Z}_p \\ \vdots & \vdots & \ddots & \vdots \\ \bar{Z}_1 & \bar{Z}_2 & \cdots & \bar{Z}_p \end{bmatrix}}$$

$$= \underset{(n\times p)}{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}} = \underset{(n\times p)}{\mathbf{0}}$$

Thus,

$$\underset{(p\times p)}{\mathbf{S_Z}} = \frac{1}{n-1} \cdot \left(\underset{(n\times p)}{\mathbf{Z}} - \frac{1}{n} \cdot \underset{(n\times 1)}{\mathbf{1}} \cdot \underset{(1\times n)}{\mathbf{1}'} \cdot \underset{(n\times p)}{\mathbf{Z}}\right)' \cdot \left(\underset{(n\times p)}{\mathbf{Z}} - \frac{1}{n} \cdot \underset{(n\times 1)}{\mathbf{1}} \cdot \underset{(1\times n)}{\mathbf{1}'} \cdot \underset{(n\times p)}{\mathbf{Z}}\right)$$

$$= \frac{1}{n-1} \cdot \left(\underset{(n\times p)}{\mathbf{Z}} - \underset{(n\times 1)}{\mathbf{1}} \cdot \underset{(1\times p)}{\bar{\mathbf{Z}}'}\right)' \cdot \left(\underset{(n\times p)}{\mathbf{Z}} - \underset{(n\times 1)}{\mathbf{1}} \cdot \underset{(1\times p)}{\bar{\mathbf{Z}}'}\right)$$

$$= \frac{1}{n-1} \cdot \left(\underset{(n\times p)}{\mathbf{Z}} - \underset{(n\times p)}{\mathbf{0}}\right)' \cdot \left(\underset{(n\times p)}{\mathbf{Z}} - \underset{(n\times p)}{\mathbf{0}}\right)$$

$$= \frac{1}{n-1} \cdot \underset{(p\times n)}{\mathbf{Z}'} \cdot \underset{(n\times p)}{\mathbf{Z}}$$

$$= \frac{1}{n-1} \cdot \underset{(p\times n)}{\begin{bmatrix} Z_{11} & Z_{21} & \cdots & Z_{n1} \\ Z_{12} & Z_{22} & \cdots & Z_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{1p} & Z_{2p} & \cdots & Z_{np} \end{bmatrix}} \underset{(n\times p)}{\begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1p} \\ Z_{21} & Z_{22} & \cdots & Z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{np} \end{bmatrix}}$$

$$
= \frac{1}{n-1}
\underset{(p\times n)}{\begin{bmatrix}
\frac{X_{11}-\bar{X}_1}{\sqrt{S_{11}}} & \frac{X_{21}-\bar{X}_1}{\sqrt{S_{11}}} & \cdots & \frac{X_{n1}-\bar{X}_1}{\sqrt{S_{11}}} \\
\frac{X_{12}-\bar{X}_2}{\sqrt{S_{22}}} & \frac{X_{22}-\bar{X}_2}{\sqrt{S_{22}}} & \cdots & \frac{X_{n2}-\bar{X}_2}{\sqrt{S_{22}}} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{X_{1p}-\bar{X}_p}{\sqrt{S_{pp}}} & \frac{X_{2p}-\bar{X}_p}{\sqrt{S_{pp}}} & \cdots & \frac{X_{np}-\bar{X}_p}{\sqrt{S_{pp}}}
\end{bmatrix}}
\underset{(n\times p)}{\begin{bmatrix}
\frac{X_{11}-\bar{X}_1}{\sqrt{S_{11}}} & \frac{X_{12}-\bar{X}_2}{\sqrt{S_{22}}} & \cdots & \frac{X_{1p}-\bar{X}_p}{\sqrt{S_{pp}}} \\
\frac{X_{21}-\bar{x}_1}{\sqrt{S_{11}}} & \frac{X_{22}-\bar{X}_2}{\sqrt{S_{22}}} & \cdots & \frac{X_{2p}-\bar{X}_p}{\sqrt{S_{pp}}} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{X_{n1}-\bar{X}_1}{\sqrt{S_{11}}} & \frac{X_{n2}-\bar{X}_2}{\sqrt{S_{22}}} & \cdots & \frac{X_{np}-\bar{X}_p}{\sqrt{S_{pp}}}
\end{bmatrix}}
$$

$$
= \underset{(p\times p)}{\begin{bmatrix}
\frac{\frac{\sum_{j=1}^{n}\left(X_{j1}-\bar{X}_1\right)^2}{\sqrt{S_{11}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{\sum_{j=1}^{n}(X_{j1}-\bar{X}_1)(X_{j2}-\bar{X}_2)}{\sqrt{S_{11}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{\sum_{j=1}^{n}(X_{j1}-\bar{X}_1)(X_{jp}-\bar{X}_p)}{\sqrt{S_{11}}\sqrt{S_{pp}}}}{n-1} \\
\frac{\frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)(X_{j1}-\bar{X}_1)}{\sqrt{S_{22}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)^2}{\sqrt{S_{22}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{\sum_{j=1}^{n}(X_{j2}-\bar{X}_2)(X_{jp}-\bar{X}_p)}{\sqrt{S_{22}}\sqrt{S_{pp}}}}{n-1} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)(X_{j1}-\bar{X}_1)}{\sqrt{S_{pp}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)(X_{j2}-\bar{X}_2)}{\sqrt{S_{pp}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{\sum_{j=1}^{n}(X_{jp}-\bar{X}_p)^2}{\sqrt{S_{pp}}\sqrt{S_{pp}}}}{n-1}
\end{bmatrix}}
$$

$$
\{S_{ik} = \tfrac{1}{n-1}\textstyle\sum_{j=1}^{n}\left(X_{ji}-\bar{X}_i\right)\left(X_{jk}-\bar{X}_k\right) \Rightarrow (n-1)S_{ik} = \sum_{j=1}^{n}\left(X_{ji}-\bar{X}_i\right)\left(X_{jk}-\bar{X}_k\right) \text{ for } i,k = 1,2,\ldots,p\}
$$

$$
= \underset{(p\times p)}{\begin{bmatrix}
\frac{\frac{(n-1)S_{11}}{\sqrt{S_{11}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{(n-1)S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{(n-1)S_{1p}}{\sqrt{S_{11}}\sqrt{S_{pp}}}}{n-1} \\
\frac{\frac{(n-1)S_{21}}{\sqrt{S_{22}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{(n-1)S_{22}}{\sqrt{S_{22}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{(n-1)S_{2p}}{\sqrt{S_{22}}\sqrt{S_{pp}}}}{n-1} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\frac{(n-1)S_{p1}}{\sqrt{S_{pp}}\sqrt{S_{11}}}}{n-1} & \frac{\frac{(n-1)S_{p2}}{\sqrt{S_{pp}}\sqrt{S_{22}}}}{n-1} & \cdots & \frac{\frac{(n-1)S_{pp}}{\sqrt{S_{pp}}\sqrt{S_{pp}}}}{n-1}
\end{bmatrix}}
$$

$$= \begin{bmatrix} \frac{S_{11}}{\sqrt{S_{11}}\sqrt{S_{11}}} & \frac{S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}} & \cdots & \frac{S_{1p}}{\sqrt{S_{11}}\sqrt{S_{pp}}} \\ \frac{S_{21}}{\sqrt{S_{22}}\sqrt{S_{11}}} & \frac{S_{22}}{\sqrt{S_{22}}\sqrt{S_{22}}} & \cdots & \frac{S_{2p}}{\sqrt{S_{22}}\sqrt{S_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{S_{p1}}{\sqrt{S_{pp}}\sqrt{S_{11}}} & \frac{S_{p2}}{\sqrt{S_{pp}}\sqrt{S_{22}}} & \cdots & \frac{S_{pp}}{\sqrt{S_{pp}}\sqrt{S_{pp}}} \end{bmatrix}_{(p\times p)}$$

$$= \begin{bmatrix} 1 & R_{12} & \cdots & R_{1p} \\ R_{21} & 1 & \cdots & R_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p1} & R_{p2} & \cdots & 1 \end{bmatrix}_{(p\times p)}$$

$$= \underset{(p\times p)}{\mathbf{R}} \blacksquare$$

Thus, the sample variance-covariance matrix $\underset{(p\times p)}{\mathbf{S_Z}}$ derived from matrix $\underset{(n\times p)}{\mathbf{Z}}$, is equivalent to the sample correlation matrix $\underset{(p\times p)}{\mathbf{R}}$, derived from $\underset{(n\times p)}{\mathbf{X}}$. That is, $\underset{(p\times p)}{\mathbf{S_Z}} = \underset{(p\times p)}{\mathbf{R}}$ .The diagonal elements of $\underset{(p\times p)}{\mathbf{S_Z}} = \underset{(p\times p)}{\mathbf{R}}$ are the sample variances

$$S_{z,kk} = \frac{S_{kk}}{\sqrt{S_{kk}}\sqrt{S_{kk}}} = \frac{S_{kk}}{S_{kk}} = R_{kk} = 1$$

for $k = 1,2, \dots, p, i = k$ where $S_{z,ii} = S_{z,kk}$. The off-diagonal elements of $\underset{(p\times p)}{\mathbf{S_Z}} = \underset{(p\times p)}{\mathbf{R}}$ are the sample covariances

$$S_{z,ik} = \frac{S_{ik}}{\sqrt{S_{ii}}\sqrt{S_{kk}}} = R_{ik}$$

for $i, k = 1,2, \dots, p, i \neq k$ where $S_{z,ik} = S_{z,ki}$.

Furthermore,

$$\operatorname{tr}(\mathbf{R}) = \sum_{k=1}^{p} R_{kk} = \operatorname{tr}(\mathbf{S}_{\mathbf{Z}}) = \sum_{k=1}^{p} S_{z,kk} = 1 + 1 + \cdots + 1 = p$$

(total standardized sample variance).

4.9 Sample Mean Vector and Variance-Covariance Matrix for Linear Combinations of Standardized Samples

4.9.1 Linear Combination of Standardized Samples

Definition 4.9.1 (Linear Combination of **Z**). *Let* $\underset{(p\times 1)}{\mathbf{c}}$ *be a* $p \times 1$ *vector of constants defined as*

$$\underset{(p\times 1)}{\mathbf{c}} = \underset{(p\times 1)}{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}}$$

and let $\underset{(p\times 1)}{\mathbf{Z}}$ *be a* $p \times 1$ *population random vector of standardized continuous random variables*

$$\underset{(p\times 1)}{\mathbf{Z}} = \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}}$$

Now consider a **linear combination of** $\underset{(n\times p)}{\mathbf{Z}}$ *with form*

$$\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = \underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \cdot \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}} = c_1 Z_1 + c_2 Z_2 + \cdots + c_p Z_p$$

whose unrealized quantity on the jth *standardized multivariate sample observation* is

$$\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}_j} = \underset{(1\times p)}{[c_1, c_2, \dots, c_p]} \cdot \underset{(p\times 1)}{\begin{bmatrix} Z_{j1} \\ Z_{j2} \\ \vdots \\ Z_{jp} \end{bmatrix}} = c_1 Z_{j1} + c_2 Z_{j2} + \cdots + c_p Z_{jp}$$

for $j = 1,2, \dots, n$.

4.9.2 Sample Statistics for Linear Combinations of Standardized Samples

Definition 4.9.2 (Sample Mean for a Linear Combination of **Z**). *Let* $\underset{(p\times 1)}{\mathbf{Z}_1}, \underset{(p\times 1)}{\mathbf{Z}_2}, \dots, \underset{(p\times 1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample defined in Theorem* 4.6.1. *Assume the sample mean vector* $\underset{(p\times 1)}{\bar{\mathbf{Z}}}$ *defined in Theorem* 4.8.1 *exists. Next, consider a linear combination of the form* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$ *with* jth *standardized multivariate sample observation* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}_j}$ *given in Definition* 4.9.1. *Then the unrealized* **sample mean for a linear combination of** $\underset{(n\times p)}{\mathbf{Z}}$ *is defined by*

$$\textit{sample mean of } \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = E\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right) = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\bar{\mathbf{Z}}} = 0.$$

Definition 4.9.3 (Sample Variance for a Linear Combination of **Z**). *Let* $\underset{(p\times 1)}{\mathbf{Z}_1}, \underset{(p\times 1)}{\mathbf{Z}_2}, \dots, \underset{(p\times 1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample defined in Theorem* 4.6.1. *Assume the sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} = \underset{(p\times p)}{\mathbf{R}}$ *defined in Theorem* 4.8.2 *exists. Next, consider a linear combination of the form* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$ *with* jth *standardized multivariate sample observation* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}_j}$ *given in Definition* 4.9.1. *Then the unrealized* **sample variance for a linear combination of** $\underset{(n\times p)}{\mathbf{Z}}$ *is defined by*

$$\textit{sample variance of } \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$$

$$= \operatorname{var}\left(\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right) = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}} = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}}.$$

Definition 4.9.4 (Sample Covariance for Two Linear Combinations of **Z**). *Let* $\underset{(p\times 1)}{\mathbf{Z}_1}, \underset{(p\times 1)}{\mathbf{Z}_2}, \dots, \underset{(p\times 1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample defined in Theorem* 4.6.1. *Assume the sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} = \underset{(p\times p)}{\mathbf{R}}$ *defined in Theorem* 4.8.2 *exists. Next, consider two linear combinations of the form* $\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$ *and* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$ *with* jth standardized *multivariate sample observations* $\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{Z}_j}$ *and* $\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}_j}$*, respectively, given in*

Definition 4.9.1. *Then the unrealized* **sample covariance for two linear combinations of** $\underset{(n\times p)}{\mathbf{Z}}$ *is defined by*

$$\textit{sample covariance of } \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{Z}} \textit{ and } \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$$

$$\operatorname{cov}\left(\underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}, \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right)$$

$$= \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}} = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{b}}$$

$$= \underset{(1\times p)}{\mathbf{b}'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}} = \underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{b}}.$$

4.9.3 q Linear Combinations of Standardized Samples

Definition 4.9.5 (q Linear Combinations of $\mathbf{Z}$). *Let* $\underset{(p\times 1)}{\mathbf{Z}_1}, \underset{(p\times 1)}{\mathbf{Z}_2}, \dots, \underset{(p\times 1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample defined in Theorem* 4.6.1. *Now consider* **q linear combinations of** $\underset{(n\times p)}{\mathbf{Z}}$ *of the form:*

$$Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = \underset{(1\times p)}{\left[c_{i1}, c_{i2}, \dots, c_{ip}\right]} \cdot \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}} = c_{i1}Z_1 + c_{i2}Z_2 + \cdots + c_{ip}Z_p$$

for $i = 1,2,\dots,q$ *linear combinations*

$$Y_1 = \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = c_{11}Z_1 + c_{12}Z_2 + \cdots + c_{1p}Z_p$$

$$Y_2 = \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = c_{21}Z_1 + c_{22}Z_2 + \cdots + c_{2p}Z_p$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$Y_q = \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = c_{q1}Z_1 + c_{q2}Z_2 + \cdots + c_{qp}Z_p$$

or in matrix notation,

$$\underset{(q\times 1)}{\mathbf{Y}} = \underset{(q\times 1)}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{bmatrix}} = \underset{(q\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{Z}} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{Z}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times 1)}{\mathbf{Z}} \end{bmatrix}} = \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}} = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{Z}}$$

where the unrealized quantity on the jth *standardized multivariate sample observation,* $j = 1,2,\dots,n$, *and* ith *linear combination,* $i = 1,2,\dots,q$, *is*

$$Y_{ji} = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}_j} = \underset{(1\times p)}{\left[c_{i1}, c_{i2}, \dots, c_{ip}\right]} \cdot \underset{(p\times 1)}{\begin{bmatrix} Z_{j1} \\ Z_{j2} \\ \vdots \\ Z_{jp} \end{bmatrix}} = c_{i1}Z_{j1} + c_{i2}Z_{j2} + \cdots + c_{ip}Z_{jp}.$$

4.9.4 Sample Mean Vector for q Linear Combinations of Standardized Samples

Definition 4.9.6 (Sample Mean Vector for q Linear Combinations of $\mathbf{Z}$). *Let* $\underset{(p\times 1)}{\mathbf{Z}_1}, \underset{(p\times 1)}{\mathbf{Z}_2}, \dots, \underset{(p\times 1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample defined in Theorem* 4.6.1. *Assume the sample mean vector* $\underset{(p\times 1)}{\bar{\mathbf{Z}}} = \underset{(p\times 1)}{\mathbf{0}}$ *defined* *in Theorem* 4.8.1 *exists. Next, consider* q *linear combinations of* $\underset{(n\times p)}{\mathbf{Z}}$ *of the form*

$Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$ *with* jth *standardized multivariate sample observation*

$Y_{ji} = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}_j}$ *given in Definition* 4.9.5. *Then the unrealized* **sample mean vector for q linear combinations of** $\underset{(n\times p)}{\mathbf{Z}}$ *is defined by*

sample mean vector of $\underset{(q\times 1)}{\mathbf{Y}}$

$$\underset{(q\times 1)}{\overline{\mathbf{Y}}} = E\left(\underset{(q\times 1)}{\mathbf{Y}}\right) = E\left(\underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right) = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\overline{\mathbf{Z}}} = \underset{(q\times 1)}{\mathbf{0}}.$$

Thus, the ith row of $\underset{(q\times 1)}{\mathbf{Y}}$ has unrealized sample mean

$$\overline{Y}_i = E(Y_i) = E\left(\underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}}\right) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\overline{\mathbf{Z}}} = 0$$

for $i = 1,2,\dots,q$.

4.9.5 Sample Variance-Covariance Matrix for q Linear Combinations of Standardized Samples

Definition 4.9.7 (Sample Variance-Covariance Matrix for q Linear Combinations of **Z**). *Let* $\underset{(p\times 1)}{\mathbf{Z}_1}, \underset{(p\times 1)}{\mathbf{Z}_2}, \dots, \underset{(p\times 1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample defined in Theorem* 4.6.1. *Assume the sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} = \underset{(p\times p)}{\mathbf{R}}$ *defined in Theorem 4.8.2 exists. Next, consider* q *linear combinations of* $\underset{(n\times p)}{\mathbf{Z}}$ *of the form* $Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$ *with* jth *standardized multivariate sample* *observation* $Y_{ji} = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}_j}$ *given in Definition* 4.9.5. *Then the unrealized symmetric* **standardized sample variance-covariance matrix for q linear combinations of** $\underset{(n\times p)}{\mathbf{Z}}$ *is defined by*

$$\underset{(q\times q)}{\mathbf{S_Y}} = \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times q)}{\mathbf{C}'}$$

$$= \underset{(q\times q)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \end{bmatrix}}$$

$$= \underset{(q\times p)}{\mathbf{C}} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times q)}{\mathbf{C}'}$$

$$= \underset{(q\times q)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_1} & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_2} & \cdots & \underset{(1\times p)}{\mathbf{c}_q'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_q} \end{bmatrix}}$$

Thus, the ith row of $\underset{(q\times 1)}{\mathbf{Y}}$ has unrealized sample variance

$$\text{var}(Y_i) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\mathbf{c}_i} = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_i}$$

for $i = 1,2,\dots,q$.

And, the ith row and kth row of $\underset{(q\times 1)}{\mathbf{Y}}$ has unrealized sample covariance

$$\operatorname{cov}(Y_i, Y_k)$$

$$= \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_k} = \underset{(1\times p)}{\mathbf{c}_k'} \cdot \underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{c}_i}$$

$$= \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_k} = \underset{(1\times p)}{\mathbf{c}_k'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\mathbf{c}_i}$$

for $i, k = 1,2, \ldots, q$.

Chapter 5

Principal Components Analysis

5.1 Introduction

A principal components analysis is concerned with explaining the variance-covariance (or correlation) structure of a set of variables through a few linear combinations of these variables. Its general objectives are (1) data reduction and (2) interpretation [3, p. 430].

5.2 Population Principal Components

Algebraically, population principal components are particular linear combinations of the p population continuous random variables $X_1, X_2, \dots, X_p$. Geometrically, these linear combinations represent the selection of a new coordinate system obtained by rotating the original system with $X_1, X_2, \dots, X_p$ as the coordinate axes. The new axes represent the directions with maximum variability and provide a simpler and more parsimonious description of the covariance (or correlation) structure.

As we shall see, principal components depend solely on the covariance matrix $\underset{(p\times p)}{\sum_{\mathbf{X}}}$ or the correlation matrix $\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$. Their development does *not* require a multivariate normal assumption. On the other hand, principal components derived for multivariate normal populations have useful interpretations in terms of the constant density ellipsoids. Further, inferences can be made from the sample components when the population is multivariate normal [3, pp. 430-431].

Let

$$\underset{(p\times 1)}{\mathbf{X}} = \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}}$$

be a population random vector for continuous random variables defined in Definition 3.2.1. Assume the corresponding population variance-covariance matrix

$$\underset{(p\times p)}{\Sigma_{\mathbf{X}}} = \underset{(p\times p)}{\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}}$$

defined in Theorem 3.2.1 is positive definite with eigenvalue and normalized-eigenvector pairs

$$\left(\lambda_1, \underset{(p\times 1)}{\mathbf{e}_1}\right), \left(\lambda_2, \underset{(p\times 1)}{\mathbf{e}_2}\right), \dots, \left(\lambda_i, \underset{(p\times 1)}{\mathbf{e}_i}\right), \dots, \left(\lambda_p, \underset{(p\times 1)}{\mathbf{e}_p}\right)$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$. That is, the λ_i's are positive and distinct. One should be aware that a population variance-covariance matrix is in general positive semi-definite [6, p. 200]. But some books still assume positive definite population variance-covariance matrix in their treatment of PCA [6, p. 206]. The purpose of the assumption here is due to the fact that the proof for Theorem 5.2.1 (ith Population Principal Component) uses Theorem 2.2.1 (Maximization of Quadratic Forms for

Points on the Unit Sphere) where $\underset{(p\times p)}{\mathbf{B}} = \underset{(p\times p)}{\Sigma_{\mathbf{X}}}$ is positive definite. To clarify, we assume $\lambda_i > 0, i = 1, \dots, p$ based on $\underset{(p\times p)}{\Sigma_{\mathbf{X}}}$ being positive definite. However, the assumption of the λ_i's being distinct ensures the $\mathbf{e}_i$'s are mutually orthogonal. In general, the λ_i's can be repeated but then the associated eigenvectors need to be chosen to be orthogonal [3, p. 432].

Therefore, for the remainder of the paper we will assume all populations variance-covariance and correlation matrices will be positive definite and the λ_i's are positive and distinct, including for the sample cases. It is our belief that these assumptions do not detract from the general concept of principal components analysis and are also seen quite often in applications. Our rationalization comes from the fact that, the variance-covariance matrix of a multivariate probability distribution is positive definite unless one variable is an *exact* linear function of the others [7].

Moving on, let the orthogonal matrix with columns being the normalized eigenvectors be

$$\underset{(p\times p)}{\mathbf{E}} = \underset{(p\times p)}{\begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix}} = \underset{(p\times p)}{\left[\underset{(p\times 1)}{\mathbf{e}_1} \vdots \underset{(p\times 1)}{\mathbf{e}_2} \vdots \cdots \vdots \underset{(p\times 1)}{\mathbf{e}_p} \right]}$$

given in Definition 2.2.18. From Definition 3.3.2 consider $q = p$ linear combinations, Y_i, of the p population continuous random variables $X_1, X_2, \dots, X_p$ with arbitrary coefficients:

$$Y_1 = \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}} = c_{11}X_1 + c_{12}X_2 + \cdots + c_{1p}X_p$$

$$Y_2 = \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}} = c_{21}X_1 + c_{22}X_2 + \cdots + c_{2p}X_p$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$Y_p = \underset{(1\times p)}{\mathbf{c}_p'} \cdot \underset{(p\times 1)}{\mathbf{X}} = c_{p1}X_1 + c_{p2}X_2 + \cdots + c_{pp}X_p$$

or in matrix notation,

$$\underset{(p\times 1)}{\mathbf{Y}} = \underset{(p\times 1)}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}} \\ \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{c}_p'} \cdot \underset{(p\times 1)}{\mathbf{X}} \end{bmatrix}} = \underset{(p\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = \underset{(p\times p)}{\mathbf{C}} \cdot \underset{(p\times 1)}{\mathbf{X}}$$

Using Theorem 3.3.5, we obtain

$$\mathrm{Var}(Y_i) = \mathrm{Var}\left(\underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_i}$$

for $i = 1,2,\ldots,p$ and

$$\mathrm{Cov}(Y_i, Y_k) = \mathrm{Cov}\left(\underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}, \underset{(1\times p)}{\mathbf{c}_k'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_k}$$

for $i, k = 1,2,\ldots,p, i \neq k$.

The population principal components are those uncorrelated linear combinations $Y_1, Y_2, \ldots, Y_p$ whose population variances are as large as possible [3, p. 431].

The first population principal component is the linear combination with maximum variance among all linear combinations. That is, it maximizes $\mathrm{Var}(Y_1) = \mathrm{Var}\left(\underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}_1}$. It is clear that $\mathrm{Var}(Y_1)$ can be increased by multiplying any $\underset{(p\times 1)}{\mathbf{c}_1}$ by some constant. To eliminate this indeterminacy, it is convenient to restrict attention to coefficient vectors of unit length. We therefore define

$$\begin{array}{c}\text{First population}\\\text{principal component}\end{array} = \text{linear combination } Y_1 = \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}} \text{ that maximizes}$$

$$\mathrm{Var}\left(\underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) \text{ subject to } \underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{c}_1} = 1$$

$$\begin{array}{c}\text{Second population}\\\text{principal component}\end{array} = \text{linear combination } Y_2 = \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}} \text{ that maximizes}$$

$$\mathrm{Var}\left(\underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) \text{ subject to } \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{c}_2} = 1 \text{ and}$$

$$\mathrm{Cov}\left(\underset{(1\times p)}{\mathbf{c}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}}, \underset{(1\times p)}{\mathbf{c}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = 0$$

And the ith step,

$$\begin{array}{c}i\text{th population}\\\text{principal component}\end{array} = \text{linear combination } Y_i = \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}} \text{ that maximizes}$$

$$\mathrm{Var}\left(\underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) \text{ subject to } \underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{c}_i} = 1 \text{ and}$$

$$\mathrm{Cov}\left(\underset{(1\times p)}{\mathbf{c}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}, \underset{(1\times p)}{\mathbf{c}_k'} \cdot \underset{(p\times 1)}{\mathbf{X}}\right) = 0 \quad \text{for} \quad k < i$$

[3, p. 431].

Theorem 5.2.1 (ith Population Principal Component). *Let* $\underset{(p\times 1)}{\mathbf{X}}$ *be a population random vector for continuous random variables defined in Definition* 3.2.1 *with associated positive-definite variance-covariance matrix* $\underset{(p\times p)}{\sum_{\mathbf{X}}}$ *defined in Theorem 3.2.1. Let* $\underset{(p\times p)}{\sum_{\mathbf{X}}}$ *have eigenvalue and normalized-eigenvector pairs* $\left(\lambda_i, \underset{(p\times 1)}{\mathbf{e}_i}\right)$ $i = 1,2,\dots,p$ *where* $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$*. Then the unrealized* ***i*****th population principal component** *is given by*

$$Y_i = \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}} = \underset{(1\times p)}{\left[e_{1i}, e_{2i}, \dots, e_{pi}\right]} \cdot \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = e_{1i}X_1 + e_{2i}X_2 + \cdots + e_{pi}X_p$$

for $i = 1,2,\dots,p$, *with unrealized population variance and covariance*

$$\mathrm{Var}(Y_i) = \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{e}_i} = \lambda_i$$

for $i = 1,2,\dots,p$ *and*

$$\mathrm{Cov}(Y_i, Y_k) = \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times p)}{\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{e}_k} = 0$$

for $i, k = 1,2,\dots,p, i \neq k$ [3, p. 432].

Proof.

Using Definition 2.2.18, Theorem 3.3.5

$$\begin{aligned}\operatorname{Var}(Y_i) &= \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{e}_i} \\ &= \underset{(1\times p)}{\mathbf{e}_i'} \cdot \left(\underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{e}_i}\right) \\ &= \underset{(1\times p)}{\mathbf{e}_i'} \cdot \left(\lambda_i \cdot \underset{(p\times 1)}{\mathbf{e}_i}\right) \\ &= \lambda_i \cdot \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times 1)}{\mathbf{e}_i} \\ &= \lambda_i \cdot 1 = \lambda_i, i = 1,2, \dots, p\end{aligned}$$

Similarly,

$$\begin{aligned}\operatorname{Cov}(Y_i, Y_k) &= \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{e}_k} \\ &= \underset{(1\times p)}{\mathbf{e}_i'} \cdot \left(\underset{(p\times p)}{\textstyle\sum_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{e}_k}\right) \\ &= \underset{(1\times p)}{\mathbf{e}_i'} \cdot \left(\lambda_k \cdot \underset{(p\times 1)}{\mathbf{e}_k}\right) \\ &= \lambda_k \cdot \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times 1)}{\mathbf{e}_k} \\ &= \lambda_k \cdot 0 = 0, i, k = 1,2, \dots, p, i \neq k.\end{aligned}$$

Next, we know from the first part of Theorem 2.2.1, with $\underset{(p\times p)}{\mathbf{B}} = \underset{(p\times p)}{\Sigma_{\mathbf{X}}}$, that

$$\max_{\underset{(p\times 1)}{\mathbf{c}} \neq \underset{(p\times 1)}{\mathbf{0}}} \frac{\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}}}{\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{c}}} = \lambda_1 \qquad \left(\text{attained when } \underset{(p\times 1)}{\mathbf{c}} = \underset{(p\times 1)}{\mathbf{e}_1}\right)$$

By Definition 2.2.18 $\underset{(1\times p)}{\mathbf{e}_1'} \cdot \underset{(p\times 1)}{\mathbf{e}_1} = 1$ since the eigenvectors are normalized. Thus,

$$\max_{\underset{(p\times 1)}{\mathbf{c}} \neq \underset{(p\times 1)}{\mathbf{0}}} \frac{\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}}}{\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{c}}} = \lambda_1 = \frac{\underset{(1\times p)}{\mathbf{e}_1'} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{e}_1}}{\underset{(1\times p)}{\mathbf{e}_1'} \cdot \underset{(p\times 1)}{\mathbf{e}_1}} = \underset{(1\times p)}{\mathbf{e}_1'} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{e}_1} = \mathrm{Var}(Y_1)$$

Similarly, using the second part of Theorem 2.2.1 we get

$$\max_{\underset{(p\times 1)}{\mathbf{c}} \perp \underset{(p\times 1)}{\mathbf{e}_1}, \underset{(p\times 1)}{\mathbf{e}_2}, \dots, \underset{(p\times 1)}{\mathbf{e}_k}} \frac{\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{c}}}{\underset{(1\times p)}{\mathbf{c}'} \cdot \underset{(p\times 1)}{\mathbf{c}}} = \lambda_{k+1}, \qquad k = 1,2,\dots,p-1$$

For the choice $\underset{(p\times 1)}{\mathbf{c}} = \underset{(p\times 1)}{\mathbf{e}_{k+1}}$, with $\underset{(p\times 1)}{\mathbf{e}_{k+1}'} \cdot \underset{(p\times 1)}{\mathbf{e}_i} = 0$,

for $i = 1,2,\dots,k$ and $k = 1,2,\dots,p-1$,

$$\frac{\underset{(1\times p)}{\mathbf{e}_{k+1}'} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{e}_{k+1}}}{\underset{(1\times p)}{\mathbf{e}_{k+1}'} \cdot \underset{(p\times 1)}{\mathbf{e}_{k+1}}} = \underset{(1\times p)}{\mathbf{e}_{k+1}'} \cdot \underset{(p\times p)}{\Sigma_{\mathbf{X}}} \cdot \underset{(p\times 1)}{\mathbf{e}_{k+1}} = \mathrm{Var}(Y_{k+1}) \ \blacksquare$$

From above, the principal components are uncorrelated and have variances equal to the eigenvalues of $\underset{(p\times p)}{\Sigma_{\mathbf{X}}}$ [3, p. 432].

Thus, the population principal components, Y_i, are given by

$$Y_1 = \underset{(1\times p)}{\mathbf{e}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}} = e_{11}X_1 + e_{21}X_2 + \cdots + e_{p1}X_p$$

$$Y_2 = \underset{(1\times p)}{\mathbf{e}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}} = e_{12}X_1 + e_{22}X_2 + \cdots + e_{p2}X_p$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$Y_p = \underset{(1\times p)}{\mathbf{e}_p'} \cdot \underset{(p\times 1)}{\mathbf{X}} = e_{1p}X_1 + e_{2p}X_2 + \cdots + e_{pp}X_p$$

or in matrix notation,

$$\underset{(p\times 1)}{\mathbf{Y}} = \underset{(p\times 1)}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{e}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}} \\ \underset{(1\times p)}{\mathbf{e}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{e}_p'} \cdot \underset{(p\times 1)}{\mathbf{X}} \end{bmatrix}} = \underset{(p\times p)}{\begin{bmatrix} e_{11} & e_{21} & \cdots & e_{p1} \\ e_{12} & e_{22} & \cdots & e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{2p} & \cdots & e_{pp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = \underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times 1)}{\mathbf{X}}$$

Theorem 5.2.2 (Total Population Variance)**.** *Let* $\underset{(p\times 1)}{\mathbf{X}}$ *be a population random vector for continuous random variables defined in Definition* 3.2.1 *with associated positive-definite variance-covariance matrix* $\underset{(p\times p)}{\sum_{\mathbf{X}}}$ *defined in Theorem 3.2.1. Let* $\underset{(p\times p)}{\sum_{\mathbf{X}}}$ *have eigenvalue and normalized-eigenvector pairs* $\left(\lambda_i, \underset{(p\times 1)}{\mathbf{e}_i}\right)$ $i = 1,2,\ldots,p$ *where* $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$*. Let* $Y_1 = \underset{(1\times p)}{\mathbf{e}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}}, Y_2 = \underset{(1\times p)}{\mathbf{e}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}}, \ldots, Y_p = \underset{(1\times p)}{\mathbf{e}_p'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *be the population principal components. Then the* **total population variance**

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} \sigma_{ii} = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} \mathrm{Var}(Y_i).$$

Proof.

From Definition 2.2.14,

$$\operatorname{tr}(\Sigma_{\mathbf{X}}) = \sum_{i=1}^{p} \sigma_{ii} = \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}.$$

Using a direct result of Result 2.2.8 with $\underset{(p\times p)}{\mathbf{A}} = \underset{(p\times p)}{\Sigma_{\mathbf{X}}}$, we can write

$$\underset{(p\times p)}{\Sigma_{\mathbf{X}}} = \underset{(p\times p)}{\mathbf{E}} \cdot \underset{(p\times p)}{\mathbf{\Lambda}} \cdot \underset{(p\times p)}{\mathbf{E}'}$$

where $\underset{(p\times p)}{\mathbf{\Lambda}}$ is the diagonal matrix of eigenvalues and $\underset{(p\times p)}{\mathbf{E}}$ is the orthogonal matrix with columns being the normalized eigenvectors.

Using Result 2.2.6 (b) and orthogonality of $\underset{(p\times p)}{\mathbf{E}}$, we have

$$\operatorname{tr}(\Sigma_{\mathbf{X}}) = \operatorname{tr}(\mathbf{E} \cdot \mathbf{\Lambda} \cdot \mathbf{E}') = \operatorname{tr}(\mathbf{\Lambda} \cdot \mathbf{E}' \cdot \mathbf{E}) = \operatorname{tr}(\mathbf{\Lambda} \cdot \mathbf{I}) = \operatorname{tr}(\mathbf{\Lambda}) = \lambda_1 + \lambda_2 + \cdots + \lambda_p$$

Thus,

$$\sum_{i=1}^{p} \sigma_{ii} = \operatorname{tr}(\Sigma_{\mathbf{X}}) = \operatorname{tr}(\mathbf{\Lambda}) = \sum_{i=1}^{p} \operatorname{Var}(Y_i) \quad \blacksquare$$

Hence,

$$\begin{aligned}\text{Total population variance} &= \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} \\ &= \lambda_1 + \lambda_2 + \cdots + \lambda_p.\end{aligned}$$

Consequently,

$$\begin{pmatrix}\text{Proportion of total}\\ \text{population variance}\\ \text{due to } i\text{th population}\\ \text{principal component}\end{pmatrix} = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \qquad i = 1,2,\dots,p$$

and

$$\begin{pmatrix}\text{Proportion of total population}\\ \text{variance due to the first } k \text{ population}\\ \text{principal components}\end{pmatrix} = \frac{\sum_{i=1}^{k} \lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \qquad k < p.$$

If most (for instance, 80 to 90%) of the total population variance, for large p, can be attributed to the first one, two, or three components, then these components can "replace" the original p variables without much loss of information.

Each component of the coefficient vector $\mathbf{e}_i' = [e_{1i}, e_{2i}, \ldots, e_{ki}, \ldots, e_{pi}]$ also merits inspection. The magnitude of e_{ki} measures the importance of the kth variable to the ith principal component, irrespective of the other variables [3, pp. 432-433].

5.3 Population Principal Components for Standardized Continuous Random Variables

The population principal components derived from a standardized population random vector for continuous random variables $\underset{(p\times 1)}{\mathbf{Z}}$ may be obtained from the normalized eigenvectors of the correlation matrix $\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$. All our previous results apply, with some simplifications, since the variance of each Z_i is unity. We shall continue to use the notation Y_i to refer to the ith population principal component and $\left(\lambda_i, \underset{(p\times 1)}{\mathbf{e}_i}\right)$ for the eigenvalue and normalized-eigenvector pair from either $\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ or $\underset{(p\times p)}{\sum_{\mathbf{X}}}$. However, the $\left(\lambda_i, \underset{(p\times 1)}{\mathbf{e}_i}\right)$ derived from $\underset{(p\times p)}{\sum_{\mathbf{X}}}$ are, in general, not the same as the ones derived from $\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ [3, p. 437].

Theorem 5.3.1 (ith Population Principal Component of **Z**). *Let* $\underset{(p\times 1)}{\mathbf{Z}}$ *be a standardized population random vector for continuous random variables defined in Definition 3.4.1. with associated positive-definite variance-covariance matrix* $\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ *defined in Theorem 3.4.4. Let* $\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ *have eigenvalue and normalized-eigenvector pairs* $\left(\lambda_i, \underset{(p\times 1)}{\mathbf{e}_i}\right), i = 1,2,\dots,p$ *where* $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$.

Then the unrealized **ith population principal component of** $\underset{(p\times 1)}{\mathbf{Z}}$ *is given by*

$$Y_i = \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times p)}{\left(\mathbf{V}^{1/2}\right)^{-1}} \cdot \left(\underset{(p\times 1)}{\mathbf{X}} - \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}}\right)$$

$$= \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times p)}{\mathbf{V}^{-1/2}} \cdot \left(\underset{(p\times 1)}{\mathbf{X}} - \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}}\right)$$

$$= \underset{(1\times p)}{\left[e_{1i}, e_{2i}, \dots, e_{pi}\right]} \cdot \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}} = e_{1i}Z_1 + e_{2i}Z_2 + \cdots + e_{pi}Z_p$$

for $i = 1,2,\dots,p$ *with unrealized population variance and covariance,*

$$\mathrm{Var}(Y_i) = \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times p)}{\textstyle\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{e}_i} = \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{e}_i} = \lambda_i$$

for $i = 1,2,\dots,p$ *and*

$$\mathrm{Cov}(Y_i, Y_k) = \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times p)}{\textstyle\sum_{\mathbf{Z}}} \cdot \underset{(p\times 1)}{\mathbf{e}_k} = \underset{(1\times p)}{\mathbf{e}_i'} \cdot \underset{(p\times p)}{\boldsymbol{\rho}} \cdot \underset{(p\times 1)}{\mathbf{e}_k} = 0$$

for $i, k = 1,2,\dots,p, i \neq k$ [3, p. 437].

Proof.

Follows from Theorem 5.2.1 with $Z_1, Z_2, \dots, Z_p$ in place of $X_1, X_2, \dots, X_p$ and

$\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ in place of $\underset{(p\times p)}{\sum_{\mathbf{X}}}$ ■

Thus, the population principal components of $\underset{(p\times 1)}{\mathbf{Z}}$, Y_i, are given by

$$Y_1 = \underset{(1\times p)}{\mathbf{e}_1'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = e_{11}Z_1 + e_{21}Z_2 + \cdots + e_{p1}Z_p$$

$$Y_2 = \underset{(1\times p)}{\mathbf{e}_2'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = e_{12}Z_1 + e_{22}Z_2 + \cdots + e_{p2}Z_p$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$Y_p = \underset{(1\times p)}{\mathbf{e}_p'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = e_{1p}Z_1 + e_{2p}Z_2 + \cdots + e_{pp}Z_p$$

or in matrix notation,

$$\underset{(p\times 1)}{\mathbf{Y}} = \underset{(p\times 1)}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix}} = \underset{(p\times 1)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{e}_1'} \cdot \underset{(p\times 1)}{\mathbf{Z}} \\ \underset{(1\times p)}{\mathbf{e}_2'} \cdot \underset{(p\times 1)}{\mathbf{Z}} \\ \vdots \\ \underset{(1\times p)}{\mathbf{e}_p'} \cdot \underset{(p\times 1)}{\mathbf{Z}} \end{bmatrix}} = \underset{(p\times p)}{\begin{bmatrix} e_{11} & e_{21} & \cdots & e_{p1} \\ e_{12} & e_{22} & \cdots & e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{2p} & \cdots & e_{pp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}}$$

$$= \underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = \underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times p)}{\left(\mathbf{V}^{1/2}\right)^{-1}} \cdot \left(\underset{(p\times 1)}{\mathbf{X}} - \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}}\right)$$

$$= \underset{(p\times p)}{\mathbf{E}'} \cdot \underset{(p\times p)}{\mathbf{V}^{-1/2}} \cdot \left(\underset{(p\times 1)}{\mathbf{X}} - \underset{(p\times 1)}{\boldsymbol{\mu}_{\mathbf{X}}}\right)$$

Theorem 5.3.2 (Total Standardized Population Variance)**.** *Let* $\underset{(p\times 1)}{\mathbf{Z}}$ *be a standardized population random vector for continuous random variables defined in Definition 3.4.1. with associated positive-definite standardized variance-covariance matrix* $\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ *defined in Theorem 3.4.4. Let* $\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ *have eigenvalue and normalized-eigenvector pairs* $\left(\lambda_i, \underset{(p\times 1)}{\mathbf{e}_i}\right), i = 1,2, \dots, p$ *where* $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0.$

Let $Y_1 = \underset{(1\times p)}{\mathbf{e}_1'} \cdot \underset{(p\times 1)}{\mathbf{Z}}, Y_2 = \underset{(1\times p)}{\mathbf{e}_2'} \cdot \underset{(p\times 1)}{\mathbf{Z}}, \dots, Y_p = \underset{(1\times p)}{\mathbf{e}_p'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$ *be the population principal components of* $\underset{(p\times 1)}{\mathbf{Z}}$.

Then the **total standardized population variance**

$$\sum_{i=1}^{p} \mathrm{Var}(Z_i) = \sum_{i=1}^{p} \mathrm{Var}(Y_i) = p.$$

Proof.

Follows from Theorem 5.2.2 with $Z_1, Z_2, \dots, Z_p$ in place of $X_1, X_2, \dots, X_p$ and

$\underset{(p\times p)}{\sum_{\mathbf{Z}}} = \underset{(p\times p)}{\boldsymbol{\rho}}$ in place of $\underset{(p\times p)}{\sum_{\mathbf{X}}}$ ■

Hence,

$$\begin{aligned}\text{Total standardized population variance} &= 1 + 1 + \cdots + 1 \\ &= p \\ &= \lambda_1 + \lambda_2 + \cdots + \lambda_p.\end{aligned}$$

Consequently,

$$\begin{pmatrix}\text{Proportion of total} \\ \text{standardized population variance} \\ \text{due to } i\text{th population principal} \\ \text{component of } \underset{(p\times 1)}{\mathbf{Z}}\end{pmatrix} = \frac{\lambda_i}{p} \qquad i = 1,2,\dots,p$$

and

$$\begin{pmatrix}\text{Proportion of total standardized} \\ \text{population variance due to the first } k \\ \text{population principal components of } \underset{(p\times 1)}{\mathbf{Z}}\end{pmatrix} = \frac{\sum_{i=1}^{k} \lambda_i}{p} \qquad k < p$$

[3, p. 437].

5.4 Sample Principal Components

Theorem 5.4.1 (ith Sample Principal Component)**.** *Let random vectors*

$\underset{(p\times1)}{\mathbf{X}_1}, \underset{(p\times1)}{\mathbf{X}_2}, \dots, \underset{(p\times1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition 4.2.2 with associated positive-definite sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S}_\mathbf{X}}$ *defined in Theorem 4.4.2. Let* $\underset{(p\times p)}{\mathbf{S}_\mathbf{X}}$ *have sample eigenvalue and normalized-eigenvector pairs* $\left(\hat{\lambda}_i, \underset{(p\times1)}{\hat{\mathbf{e}}_i}\right), i = 1,2,\dots,p$ *where* $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_p > 0$. *Then the unrealized* i**th sample principal component** *is of the form*

$$\hat{Y}_i = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times1)}{\mathbf{X}} = \underset{(1\times p)}{\left[\hat{e}_{1i}, \hat{e}_{2i}, \dots, \hat{e}_{pi}\right]} \cdot \underset{(p\times1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = \hat{e}_{1i}X_1 + \hat{e}_{2i}X_2 + \cdots + \hat{e}_{pi}X_p$$

for $i = 1,2,\dots,p$ *with unrealized quantity on the* jth *multivariate sample observation*

$$\hat{Y}_{ji} = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times1)}{\mathbf{X}_j} = \underset{(1\times p)}{\left[\hat{e}_{1i}, \hat{e}_{2i}, \dots, \hat{e}_{pi}\right]} \cdot \underset{(p\times1)}{\begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jp} \end{bmatrix}} = \hat{e}_{1i}X_{j1} + \hat{e}_{2i}X_{j2} + \cdots + \hat{e}_{pi}X_{jp}$$

for $j = 1,2,\dots,n$, *with unrealized sample variance and covariance*

$$\operatorname{var}\left(\hat{Y}_i\right) = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times p)}{\mathbf{S}_\mathbf{X}} \cdot \underset{(p\times1)}{\hat{\mathbf{e}}_i} = \hat{\lambda}_i$$

for $i = 1,2,\dots,p$ *and*

$$\operatorname{cov}\left(\hat{Y}_i, \hat{Y}_k\right) = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times p)}{\mathbf{S}_\mathbf{X}} \cdot \underset{(p\times1)}{\hat{\mathbf{e}}_k} = 0$$

for $i, k = 1,2, \ldots, p, i \neq k$ *given in Definition 4.5.4* [3, p. 442].

One can write the n sample principal components in matrix notation

$$\underset{(n\times p)}{\widehat{\mathbf{Y}}} = \underset{(n\times p)}{\mathbf{X}} \cdot \underset{(p\times p)}{\widehat{\mathbf{E}}}$$

$$= \underset{(n\times p)}{\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ X_{j1} & X_{j2} & \cdots & X_{jp} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}} \underset{(p\times p)}{\begin{bmatrix} \hat{e}_{11} & \hat{e}_{12} & \cdots & \hat{e}_{1i} & \cdots & \hat{e}_{1p} \\ \hat{e}_{21} & \hat{e}_{22} & \cdots & \hat{e}_{2i} & \cdots & \hat{e}_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \hat{e}_{p1} & \hat{e}_{p2} & \cdots & \hat{e}_{pi} & \cdots & \hat{e}_{pp} \end{bmatrix}}$$

$$= \underset{(n\times p)}{\begin{bmatrix} \underset{(1\times p)}{\mathbf{X}_1'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_1} & \underset{(1\times p)}{\mathbf{X}_1'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_2} & \cdots & \underset{(1\times p)}{\mathbf{X}_1'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_i} & \cdots & \underset{(1\times p)}{\mathbf{X}_1'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_p} \\ \underset{(1\times p)}{\mathbf{X}_2'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_1} & \underset{(1\times p)}{\mathbf{X}_2'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_2} & \cdots & \underset{(1\times p)}{\mathbf{X}_2'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_i} & \cdots & \underset{(1\times p)}{\mathbf{X}_2'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \underset{(1\times p)}{\mathbf{X}_j'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_1} & \underset{(1\times p)}{\mathbf{X}_j'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_2} & \cdots & \underset{(1\times p)}{\mathbf{X}_j'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_i} & \cdots & \underset{(1\times p)}{\mathbf{X}_j'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \underset{(1\times p)}{\mathbf{X}_n'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_1} & \underset{(1\times p)}{\mathbf{X}_n'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_2} & \cdots & \underset{(1\times p)}{\mathbf{X}_n'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_i} & \cdots & \underset{(1\times p)}{\mathbf{X}_n'} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_p} \end{bmatrix}}$$

Using Definition 2.1.11 inner (dot) product of two vectors

$$\underset{(1\times n)}{\mathbf{x}'} \cdot \underset{(n\times 1)}{\mathbf{y}} = \underset{(1\times n)}{\mathbf{y}'} \cdot \underset{(n\times 1)}{\mathbf{x}} \Rightarrow$$

$$= \underset{(n\times p)}{\begin{bmatrix} \underset{(1\times p)}{\hat{\mathbf{e}}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}_1} & \underset{(1\times p)}{\hat{\mathbf{e}}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}_1} & \cdots & \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}_1} & \cdots & \underset{(1\times p)}{\hat{\mathbf{e}}_p'} \cdot \underset{(p\times 1)}{\mathbf{X}_1} \\ \underset{(1\times p)}{\hat{\mathbf{e}}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}_2} & \underset{(1\times p)}{\hat{\mathbf{e}}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}_2} & \cdots & \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}_2} & \cdots & \underset{(1\times p)}{\hat{\mathbf{e}}_p'} \cdot \underset{(p\times 1)}{\mathbf{X}_2} \\ \vdots & \vdots & & \vdots & & \vdots \\ \underset{(1\times p)}{\hat{\mathbf{e}}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} & \underset{(1\times p)}{\hat{\mathbf{e}}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} & \cdots & \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} & \cdots & \underset{(1\times p)}{\hat{\mathbf{e}}_p'} \cdot \underset{(p\times 1)}{\mathbf{X}_j} \\ \vdots & \vdots & & \vdots & & \vdots \\ \underset{(1\times p)}{\hat{\mathbf{e}}_1'} \cdot \underset{(p\times 1)}{\mathbf{X}_n} & \underset{(1\times p)}{\hat{\mathbf{e}}_2'} \cdot \underset{(p\times 1)}{\mathbf{X}_n} & \cdots & \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}_n} & \cdots & \underset{(1\times p)}{\hat{\mathbf{e}}_p'} \cdot \underset{(p\times 1)}{\mathbf{X}_n} \end{bmatrix}}$$

$$= \begin{bmatrix} \hat{Y}_{11} & \hat{Y}_{12} & \cdots & \hat{Y}_{1i} & \cdots & \hat{Y}_{1p} \\ \hat{Y}_{21} & \hat{Y}_{22} & \cdots & \hat{Y}_{2i} & \cdots & \hat{Y}_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \hat{Y}_{j1} & \hat{Y}_{j2} & \cdots & \hat{Y}_{ji} & \cdots & \hat{Y}_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ \hat{Y}_{n1} & \hat{Y}_{n2} & \cdots & \hat{Y}_{ni} & \cdots & \hat{Y}_{np} \end{bmatrix}_{(n\times p)}$$

Theorem 5.4.2 (Total Sample Variance)**.** *Let random vectors* $\underset{(p\times 1)}{\mathbf{X}_1}, \underset{(p\times 1)}{\mathbf{X}_2}, \ldots, \underset{(p\times 1)}{\mathbf{X}_n}$ *constitute a multivariate random sample defined in Definition 4.2.2. with associated positive-definite sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}}$ *defined in Theorem 4.4.2.*

Let $\underset{(p\times p)}{\mathbf{S}_{\mathbf{X}}}$ *have sample eigenvalue and normalized- eigenvector pairs* $\left(\hat{\lambda}_i, \underset{(p\times 1)}{\hat{\mathbf{e}}_i}\right)$, $i = 1,2,\ldots,p$ *where* $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_p > 0$. *Let the unrealized sample principal components be of the form* $\hat{Y}_i = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}}$ *with* j*th multivariate sample observation* $\hat{Y}_{ji} = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times 1)}{\mathbf{X}_j}$. Then the **total sample variance**

$$S_{11} + S_{22} + \cdots + S_{pp} = \sum_{i=1}^{p} S_{ii} = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p = \sum_{i=1}^{p} \mathrm{Var}(\hat{Y}_i).$$

Consequently,

$$\begin{pmatrix} \text{Proportion of total} \\ \text{sample variance} \\ \text{due to } i\text{th sample} \\ \text{principal component} \end{pmatrix} = \frac{\hat{\lambda}_i}{\hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p} \qquad i = 1,2,\ldots,p$$

[3, p. 442] and

$$\begin{pmatrix} \text{Proportion of total sample} \\ \text{variance due to the first } k \\ \text{sample principal components} \end{pmatrix} = \frac{\sum_{i=1}^{k} \hat{\lambda}_i}{\hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p} \qquad k < p.$$

We shall denote the sample principal components by $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_p$, irrespective of whether they are obtained from $\underset{(p\times p)}{\mathbf{S_X}}$ or $\underset{(p\times p)}{\mathbf{S_Z}} = \underset{(p\times p)}{\mathbf{R}}$. The components constructed from $\underset{(p\times p)}{\mathbf{S_X}}$ and $\underset{(p\times p)}{\mathbf{S_Z}} = \underset{(p\times p)}{\mathbf{R}}$ are not the same, in general, but it will be clear from the context which matrix is being used, and the single notation $\hat{Y}_i$ is convenient. It is also convenient to label the component coefficient vectors $\underset{(p\times 1)}{\hat{\mathbf{e}}_i}$ and the component $\hat{\lambda}_i$ for both situations [3, p. 443].

5.5 Sample Principal Components for Standardized Samples

Theorem 5.5.1 (ith Sample Principal Component of $\mathbf{Z}$). *Let random vectors* $\underset{(p\times 1)}{\mathbf{Z}_1}, \underset{(p\times 1)}{\mathbf{Z}_2}, \dots, \underset{(p\times 1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample defined in Theorem 4.6.1. with associated positive-definite sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S_Z}} = \underset{(p\times p)}{\mathbf{R}}$ *defined in Theorem 4.8.2. Let* $\underset{(p\times p)}{\mathbf{S_Z}} = \underset{(p\times p)}{\mathbf{R}}$ *have sample eigenvalue and normalized-eigenvector pairs* $\left(\hat{\lambda}_i, \underset{(p\times 1)}{\hat{\mathbf{e}}_i}\right), i = 1,2,\dots,p$ *where* $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_p > 0$*. Then the unrealized* ***i*****th sample principal component of** $\underset{(n\times p)}{\mathbf{Z}}$ *is of the form*

$$\hat{Y}_i = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}} = \underset{(1\times p)}{\left[\hat{e}_{1i}, \hat{e}_{2i}, \dots, \hat{e}_{pi}\right]} \cdot \underset{(p\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}} = \hat{e}_{1i}Z_1 + \hat{e}_{2i}Z_2 + \cdots + \hat{e}_{pi}Z_p$$

for $i = 1,2,\dots,p$ *with unrealized quantity on the* jth *standardized multivariate sample observation*

$$\hat{Y}_{ji} = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}_j} = \underset{(1\times p)}{\left[\hat{e}_{1i}, \hat{e}_{2i}, \dots, \hat{e}_{pi}\right]} \cdot \underset{(p\times 1)}{\begin{bmatrix} Z_{j1} \\ Z_{j2} \\ \vdots \\ Z_{jp} \end{bmatrix}} = \hat{e}_{1i}Z_{j1} + \hat{e}_{2i}Z_{j2} + \cdots + \hat{e}_{pi}Z_{jp}$$

for $j = 1,2,\dots,n$ *with unrealized sample variance and covariance*

$$\operatorname{var}(\hat{Y}_i) = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_i} = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_i} = \hat{\lambda}_i$$

for $i = 1,2,\dots,p$ *and*

$$\operatorname{cov}(\hat{Y}_i, \hat{Y}_k) = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times p)}{\mathbf{S_Z}} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_k} = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times p)}{\mathbf{R}} \cdot \underset{(p\times 1)}{\hat{\mathbf{e}}_k} = 0$$

for $i, k = 1,2,\dots,p, i \neq k$ *given in Definition 4.9.7* [3, p. 451].

One can write the n sample principal components of $\underset{(n\times p)}{\mathbf{Z}}$ in matrix notation

$$\underset{(n\times p)}{\hat{\mathbf{Y}}} = \underset{(n\times p)}{\mathbf{Z}} \cdot \underset{(p\times p)}{\hat{\mathbf{E}}}$$

$$= \underset{(n\times p)}{\begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1p} \\ Z_{21} & Z_{22} & \cdots & Z_{2p} \\ \vdots & \vdots & & \vdots \\ Z_{j1} & Z_{j2} & \cdots & Z_{jp} \\ \vdots & \vdots & & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{np} \end{bmatrix}} \underset{(p\times p)}{\begin{bmatrix} \hat{e}_{11} & \hat{e}_{12} & \cdots & \hat{e}_{1i} & \cdots & \hat{e}_{1p} \\ \hat{e}_{21} & \hat{e}_{22} & \cdots & \hat{e}_{2i} & \cdots & \hat{e}_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \hat{e}_{p1} & \hat{e}_{p2} & \cdots & \hat{e}_{pi} & \cdots & \hat{e}_{pp} \end{bmatrix}}$$

$$= \underset{(n\times p)}{\begin{bmatrix} \hat{Y}_{11} & \hat{Y}_{12} & \cdots & \hat{Y}_{1i} & \cdots & \hat{Y}_{1p} \\ \hat{Y}_{21} & \hat{Y}_{22} & \cdots & \hat{Y}_{2i} & \cdots & \hat{Y}_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \hat{Y}_{j1} & \hat{Y}_{j2} & \cdots & \hat{Y}_{ji} & \cdots & \hat{Y}_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ \hat{Y}_{n1} & \hat{Y}_{n2} & \cdots & \hat{Y}_{ni} & \cdots & \hat{Y}_{np} \end{bmatrix}}$$

Theorem 5.5.2 (Total Standardized Sample Variance)**.** *Let random vectors* $\underset{(p\times 1)}{\mathbf{Z}_1}, \underset{(p\times 1)}{\mathbf{Z}_2}, \dots, \underset{(p\times 1)}{\mathbf{Z}_n}$ *constitute a standardized multivariate random sample defined in Theorem 4.6.1 with associated positive-definite standardized sample variance-covariance matrix* $\underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} = \underset{(p\times p)}{\mathbf{R}}$ *defined in Theorem 4.8.2. Let* $\underset{(p\times p)}{\mathbf{S}_{\mathbf{Z}}} = \underset{(p\times p)}{\mathbf{R}}$ *have sample eigenvalue and normalized-eigenvector pairs* $\left(\hat{\lambda}_i, \underset{(p\times 1)}{\hat{\mathbf{e}}_i}\right), i = 1,2,\dots,p$ *where* $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_p > 0$*. Let the unrealized sample principal components of* $\underset{(n\times p)}{\mathbf{Z}}$ *be of the form* $\hat{Y}_i = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}}$ *with* j*th standardized multivariate sample observations* $\hat{Y}_{ji} = \underset{(1\times p)}{\hat{\mathbf{e}}_i'} \cdot \underset{(p\times 1)}{\mathbf{Z}_j}$.

Then the **total standardized sample variance**

$$1 + 1 + \cdots + 1 = p = \sum_{i=1}^{p} S_{z,ii} = \sum_{i=1}^{p} R_{ii} = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p = \sum_{i=1}^{p} \mathrm{Var}(\hat{Y}_i).$$

Consequently,

$$\left(\begin{array}{c}\text{Proportion of total}\\ \text{standardized sample variance}\\ \text{due to the } i\text{th sample principal}\\ \text{component of } \underset{(n\times p)}{\mathbf{Z}}\end{array}\right) = \frac{\hat{\lambda}_i}{p} \qquad i = 1,2,\dots,p$$

and

$$\left(\begin{array}{c}\text{Proportion of total standardized}\\ \text{sample variance due to the first } k\\ \text{sample principal components of } \underset{(n\times p)}{\mathbf{Z}}\end{array}\right) = \frac{\sum_{i=1}^{k}\hat{\lambda}_i}{p} \qquad k < p.$$

A rule of thumb suggests retaining only those components whose variances $\hat{\lambda}_i$ are greater than unity or, equivalently, only those components which, individually, explain at least a proportion $1/p$ of the total variance. This rule does not have a great deal of theoretical support, however, and it should not be applied blindly. Also, a scree plot is useful for selecting the appropriate number of components [3, p. 451].

Chapter 6

Results and Discussion

6.1 R Programming Language

Analysis of data is conducted using R version 3.6.2 (2019-12-12) -- "Dark and Stormy Night". R is an open source software for statistical computing and graphics. The latest version can be downloaded at R: The R Project for Statistical Computing website https://www.r-project.org/.

6.2 Univariate Distribution Analysis

6.2.1 Descriptives for US Crime 2018

Table 6.2.1: Descriptives for US Crime 2018

	vars	n	sd	min	q1	median	mean	q3	max	range
MURDER	1	327	5.45	0	1.95	3.9	5.11	6.25	60.9	60.9
RAPE	2	327	26.36	13	33.15	44.8	50.99	62.2	200.1	187.1
ROBBERY	3	327	61.93	1.2	33.3	55.5	70.52	87.95	473.2	472
ASSAULT	4	327	182.33	30.2	152.1	233.6	270.11	323.7	1477.8	1447.6
BURGLARY	5	327	233.2	87.3	264.1	393.9	435.9	557.1	1576.1	1488.8
LARCENY	6	327	734.07	488.5	1282	1657.2	1748.4	2045.1	8558.1	8069.6
VEHICLE	7	327	160.58	13.7	103	166.7	215.18	281.4	970.9	957.2

Table 6.2.1 gives the descriptives for 327 US metropolitan statistical areas in 2018 for violent crime and property crime per 100,000 residents.

6.2.2 Distributions of US Crime 2018

6.2.2.1 Murder Distribution

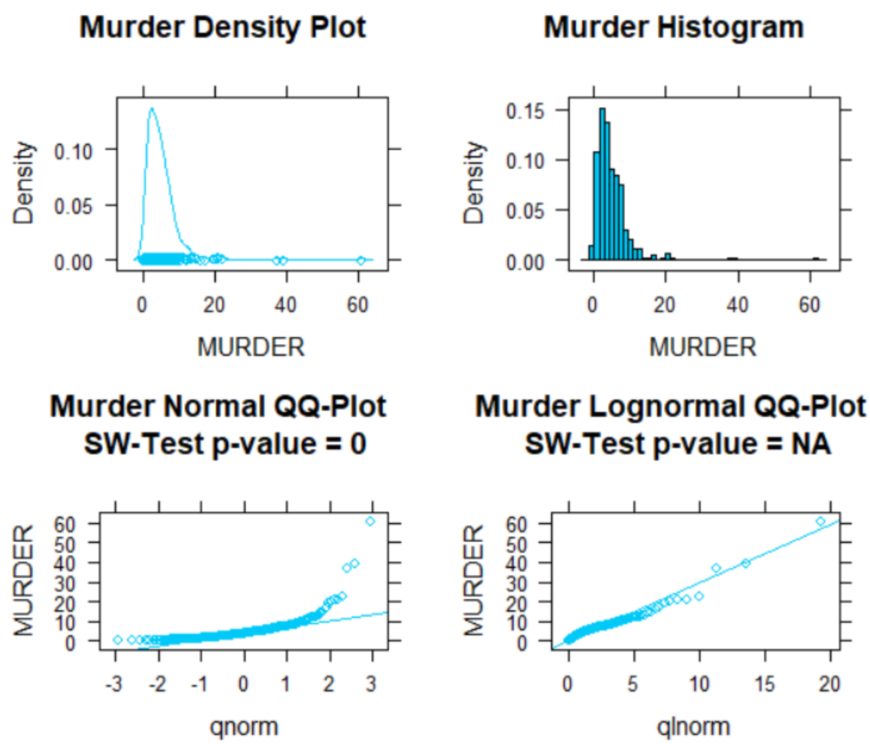


Figure 6.2.1: Murder Distribution Plots

Based on the density and histogram in Figure 6.2.1, the distribution of Murder looks right skewed. The lower left plot in Figure 6.2.1 is a Normal QQ-Plot for Murder that shows a clear lack of normality. One can use Shapiro-Wilk test for normality with $\alpha = 0.1$ to confirm this assertion. That is,

$$H_0 : \text{ Population Distribution for Murder is Normal}$$

$$H_1 : \text{ Population Distribution for Murder is not Normal}$$

$$\mathrm{W} = 0.61016; \mathrm{p} - \text{value} \cong 0$$

Thus, as expected, one rejects H_0. There is sufficient evidence to say that the population distribution of Murder is not normally distributed. However, the distribution of Murder could be lognormal. The lower right plot in Figure 6.2.1 is a Lognormal QQ-Plot for Murder that shows a clear potential of lognormality, along with the density and histogram. One can use the same Shapiro-Wilk test to test for lognormality by a simple log transformation on $\underset{(327\times 1)}{\mathbf{X}_{\text{Murder}}}$. Indeed, this is due to the fact that $X_i \sim \text{Lognormal} \Rightarrow \log(X_i) \sim \text{Normal}$ [8].

$$H_0 : \text{ Population Distribution for Murder is Lognormal}$$

$$H_1 : \text{ Population Distribution for Murder is not Lognormal}$$

$$\mathrm{p} - \text{value doesn't exist}$$

The $\mathrm{p} - \text{value}$ doesn't exist because seven metropolitan statistical areas have murder rates of 0. As a result, the transformation from $\underset{(327\times 1)}{\mathbf{X}_{\text{Murder}}}$ to $\log\left(\underset{(327\times 1)}{\mathbf{X}_{\text{Murder}}}\right)$ cannot be completed and the Shapiro-Wilk test will not compute a $\mathrm{p} - \text{value}$. Nevertheless, using the Lognormal QQ-Plot one can cautiously assume the population distribution of Murder is approximately lognormal.

It has been found that all the outliers of Murder are located at the upper end of the distribution. These metropolitan statistical areas correspond places with extremely high murder rates per 100,000 residents. Furthermore, it may be of interest to see the areas in the lowest 2.5% of the Murder distribution for 2018.

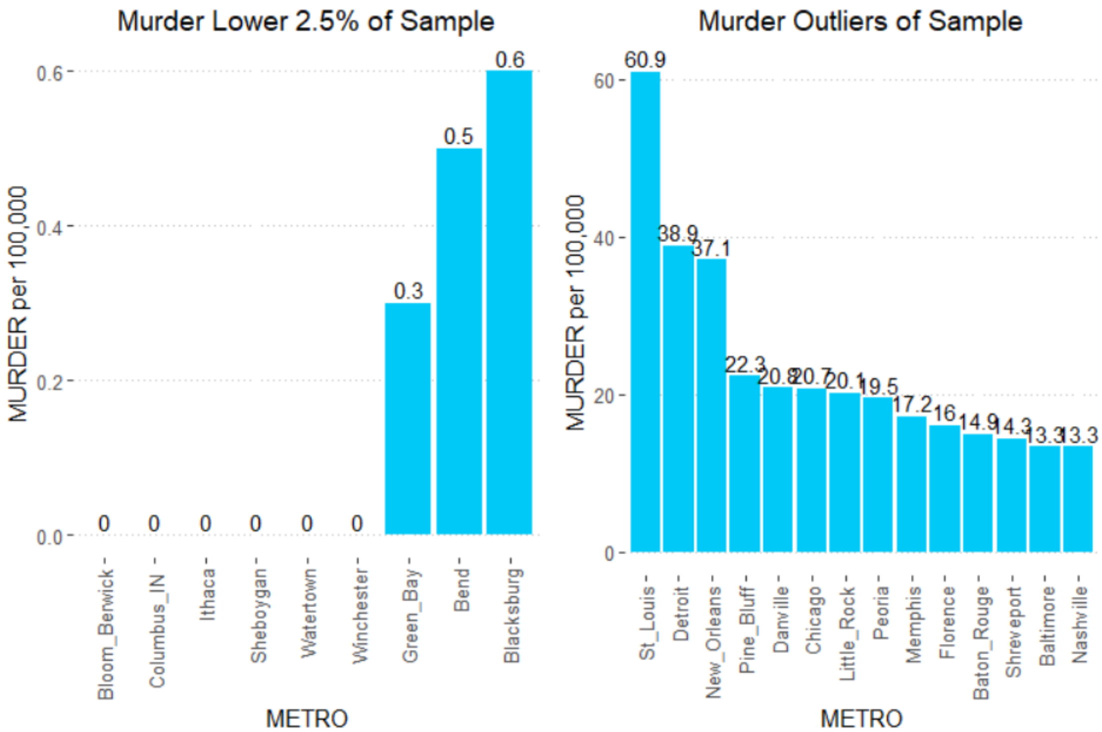


Figure 6.2.2: Murder Outliers and Lower 2.5% of Sample

The left plot in Figure 6.2.2 shows the seven metropolitan statistical areas with a murder rate of 0. The right plot in Figure 6.2.2 highlights three areas with radically high murder rates per 100,000; namely, St Louis (60.9), Detroit (38.9), and New Orleans (37.1).

6.2.2.2 Rape Distribution

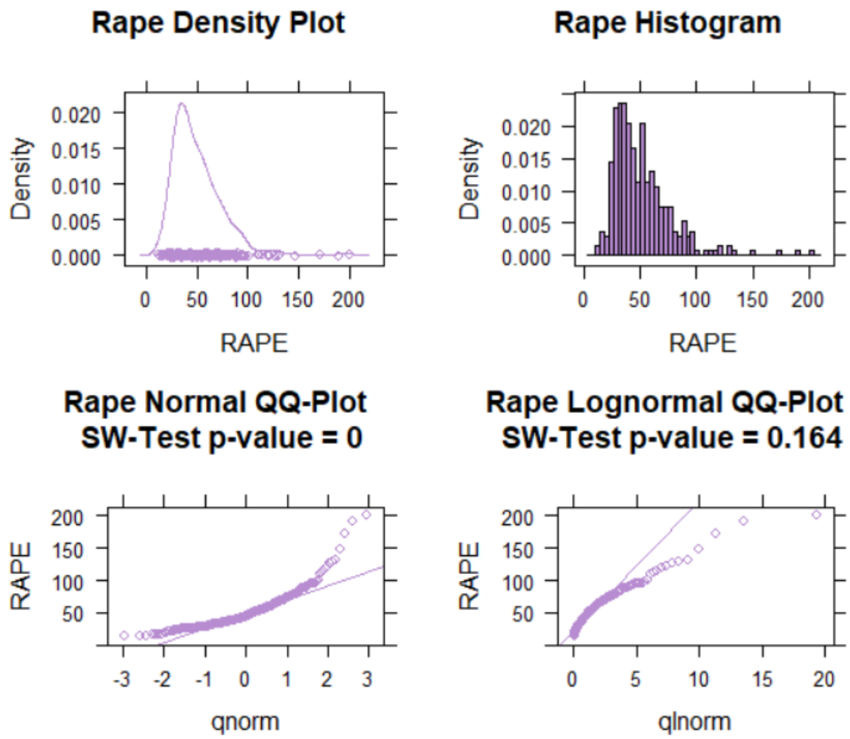


Figure 6.2.3: Rape Distribution Plots

Based on the density and histogram in Figure 6.2.3, the distribution of rape looks right skewed with several outliers. Next, one uses Shapiro-Wilk test to test for normality and lognormality with $\alpha = 0.1$.

$$H_0 : \text{ Population Distribution for Rape is Normal}$$

$$H_1 : \text{ Population Distribution for Rape is not Normal}$$

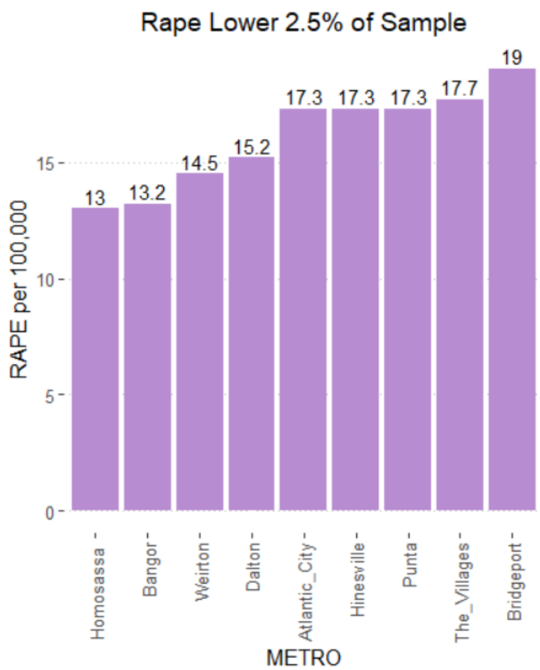
$$\mathrm{W} = 0.85053; \mathrm{p-value} \cong 0$$

$$H_0: \text{ Population Distribution for Rape is Lognormal}$$

$$H_1: \text{ Population Distribution for Rape is not Lognormal}$$

$$\text{p} - \text{value} \cong 0.164$$

One rejects H_0 for normality and fails to reject H_0 for lognormality. Yet the Lognormal QQ-Plot appears to contradict the hypothesis test result. Thus, more work should be done to resolve this inconsistency. However, learning the true distribution of rape is not of major interest, so one can move on.



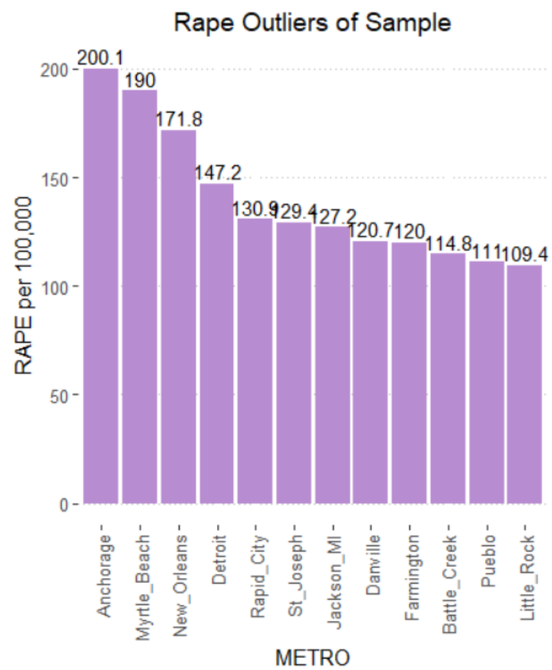


Figure 6.2.4: Rape Outliers and Lower 2.5% of Sample

The right plot in Figure 6.2.4 focuses one's attention to four areas with extremely high rape rates per 100,000; specifically, Anchorage (200.1), Myrtle Beach (190), New Orleans (171.8), and Detroit (147.2). Anchorage has long time been known for its high rape rates. The question of interest is why? Some have posed that it is related to the high male-to-female ratio. Others have said it is due to the long winters and physical isolation of individuals. While others have stated that the issue is established upon patriarchy and capitalism, which objectifies and commodifies women as the property of men [9]. Whereas, Myrtle Beach and New Orleans are vacation and party destinations which could lead to increased sexual assault. Finally, remember, that Detroit and New Orleans also had dangerously high Murder rates. One should pay attention to these metropolitan statistical areas that repeatedly show up in the high-ranking crime category.

6.2.2.3 Robbery Distribution

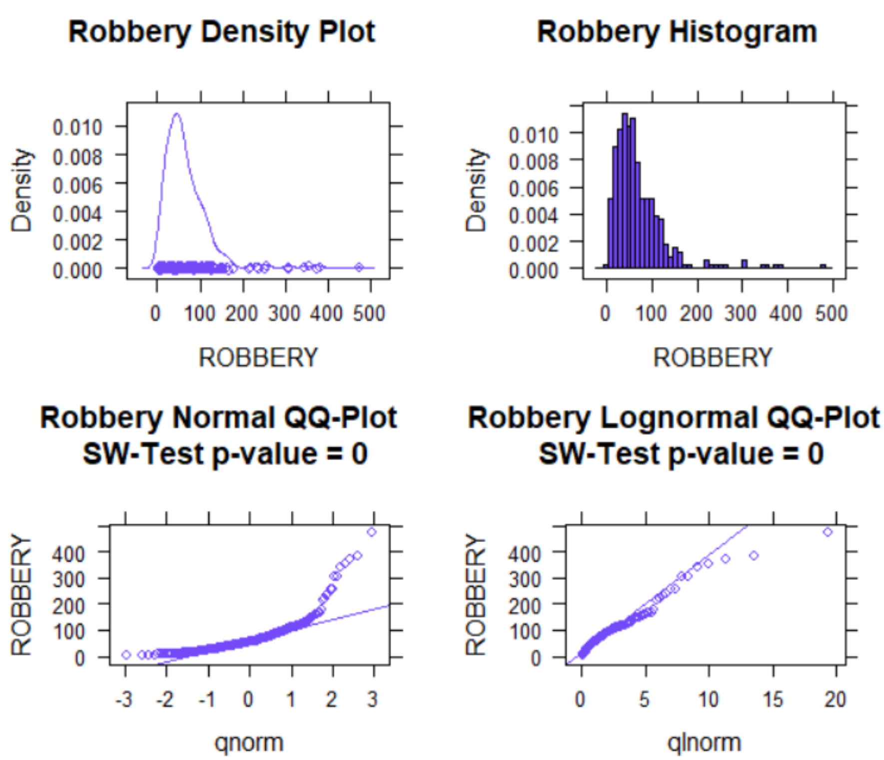


Figure 6.2.5: Robbery Distribution Plots

Based on the density and histogram in Figure 6.2.5, the distribution of Robbery looks right skewed with several outliers. Next, one uses Shapiro-Wilk test for testing normality and lognormality with $\alpha = 0.1$.

$$H_0 : \text{ Population Distribution for Robbery is Normal}$$

$$H_1 : \text{ Population Distribution for Robbery is not Normal}$$

$$\mathrm{W = 0.73871; p - value \cong 0}$$

$$H_0 : \text{ Population Distribution for Robbery is Lognormal}$$

$$H_1 : \text{ Population Distribution for Robbery is not Lognormal}$$

$$\mathrm{p - value \cong 0}$$

One rejects H_0 for normality and lognormality.

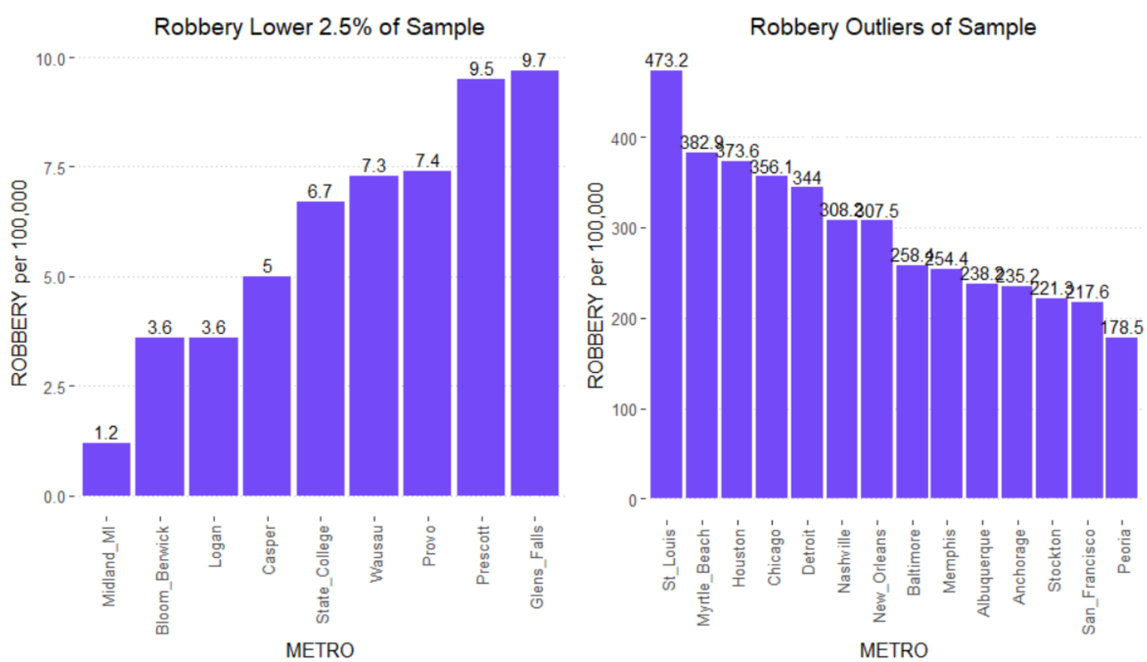


Figure 6.2.6: Robbery Outliers and Lower 2.5% of Sample

The right plot in Figure 6.2.6 has some repeatedly high-ranking metropolitan statistical areas for crime, in general, and in robbery as well. The names one hasn't seen yet in the univariate outliers list are Houston, Albuquerque, Stockton, and San Francisco.

6.2.2.4 Assault Distribution

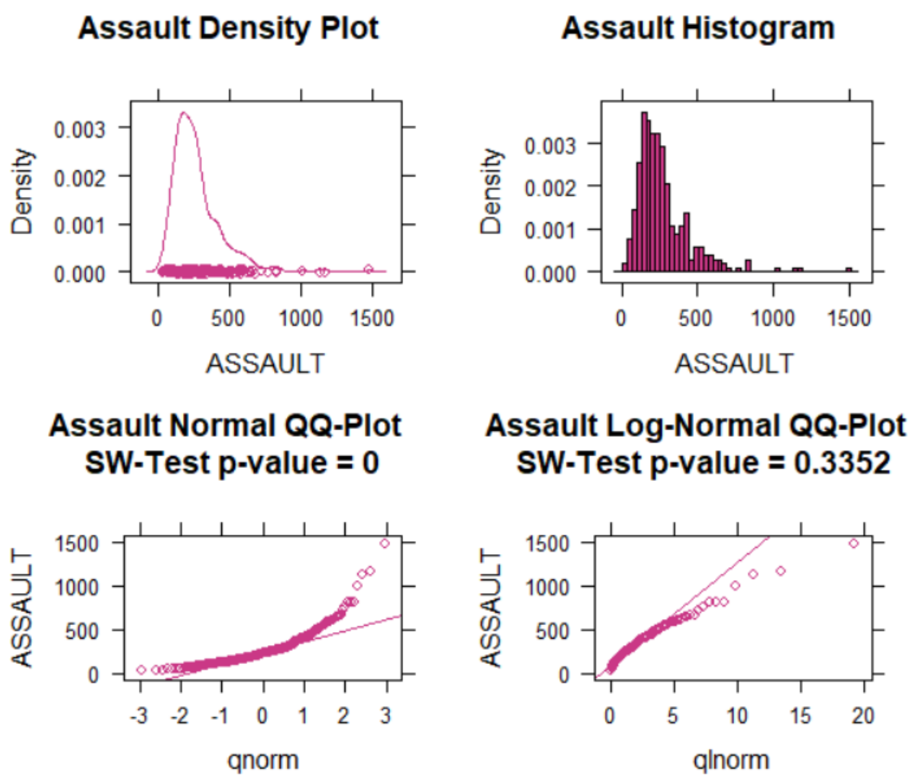


Figure 6.2.7: Assault Distribution Plots

Based on the density and histogram in Figure 6.2.7, the distribution of Assault looks right skewed with several outliers. Next, one uses Shapiro-Wilk test for normality and lognormality with $\alpha = 0.1$.

$$H_0: \text{ Population Distribution for Assault is Normal}$$

$$H_1: \text{ Population Distribution for Assault is not Normal}$$

$$\mathrm{W} = 0.8109; \mathrm{p-value} \cong 0$$

$$H_0: \text{ Population Distribution for Assault is Lognormal}$$

$$H_1: \text{ Population Distribution for Assault is not Lognormal}$$

$$\mathrm{p-value} \cong 0.3352$$

One rejects H_0 for normality and fails to reject H_0 for lognormality. Similar to rape, the Lognormal QQ-Plot for assault, appears to contradict the hypothesis test result.

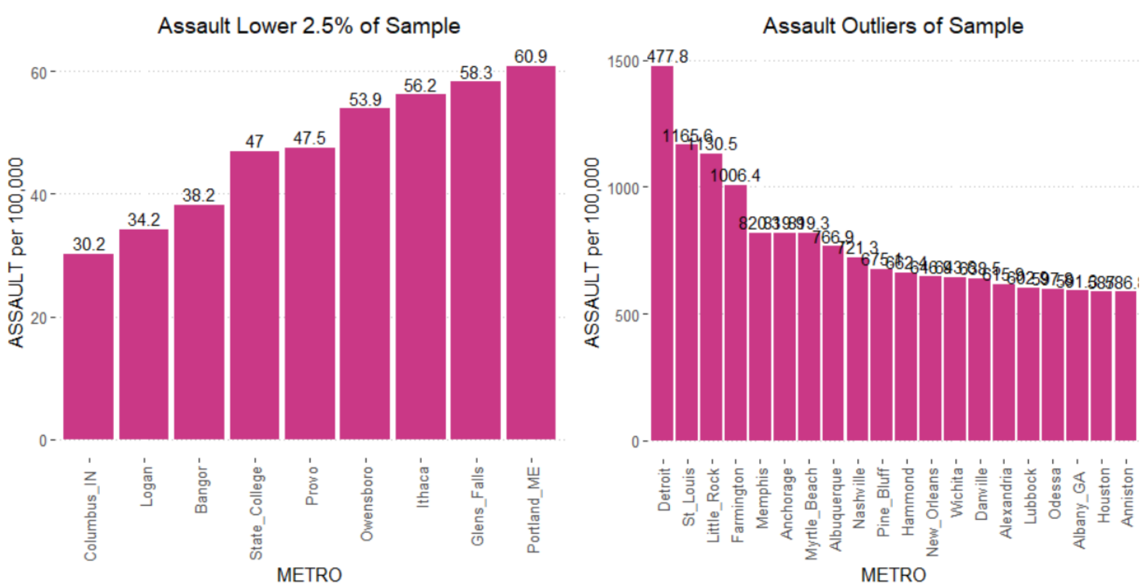


Figure 6.2.8: Assault Outliers and Lower 2.5% of Sample

The right plot in Figure 6.2.8 features four areas with drastically higher assault rates per 100,000. Detroit (1477.8), St Louis (1165.6), Little Rock (1130.5), and Farmington (1006.4). Interesting, two metropolitan statistical areas are in New Mexico: Farmington and Albuquerque. Similarly, three metropolitan statistical areas are in Texas: Lubbock, Odessa, and Houston. Immediately we can see many of these outliers have been seen in previous plots.

6.2.2.5 Burglary Distribution

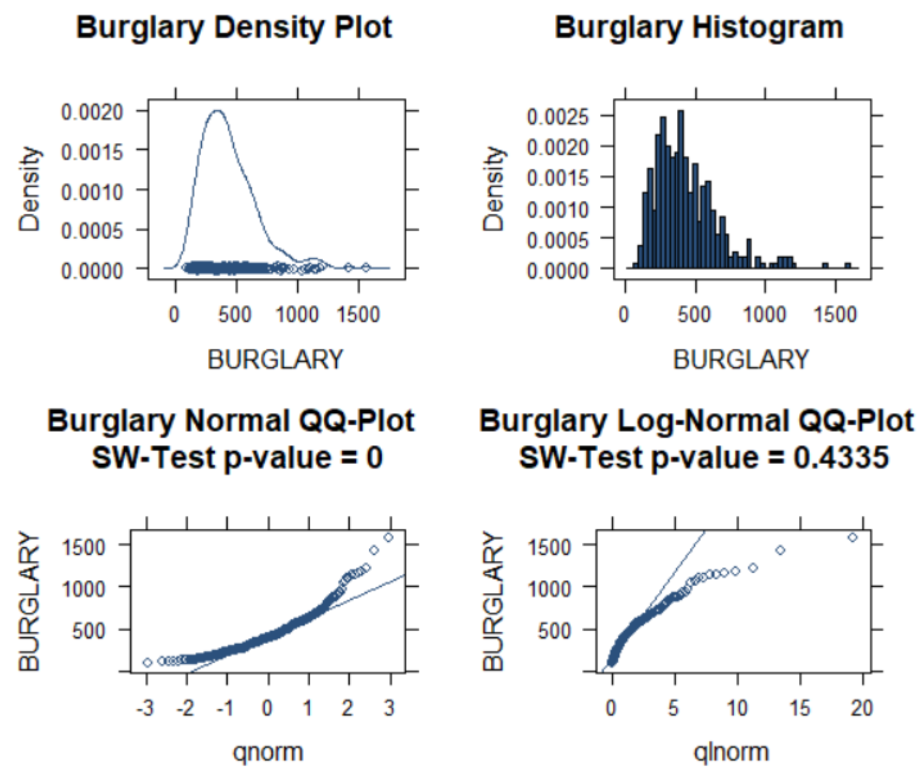


Figure 6.2.9: Burglary Distribution Plots

Based on the density and histogram in Figure 6.2.9, the distribution of Burglary looks right skewed with several outliers. Next, one uses Shapiro-Wilk test for normality and lognormality with $\alpha = 0.1$.

$$H_0:\ \text{Population Distribution for Burglary is Normal}$$

$$H_1:\ \text{Population Distribution for Burglary is not Normal}$$

$$\mathrm{W} = 0.90035;\ \mathrm{p-value} \cong 0$$

$$H_0:\ \text{Population Distribution for Burglary is Lognormal}$$

$$H_1:\ \text{Population Distribution for Burglary is not Lognormal}$$

$$\mathrm{p-value} \cong 0.4335$$

One rejects H_0 for normality and fails to reject H_0 for lognormality. Similar to rape and assault, the Lognormal QQ-Plot for burglary, appears to contradict the hypothesis test result.

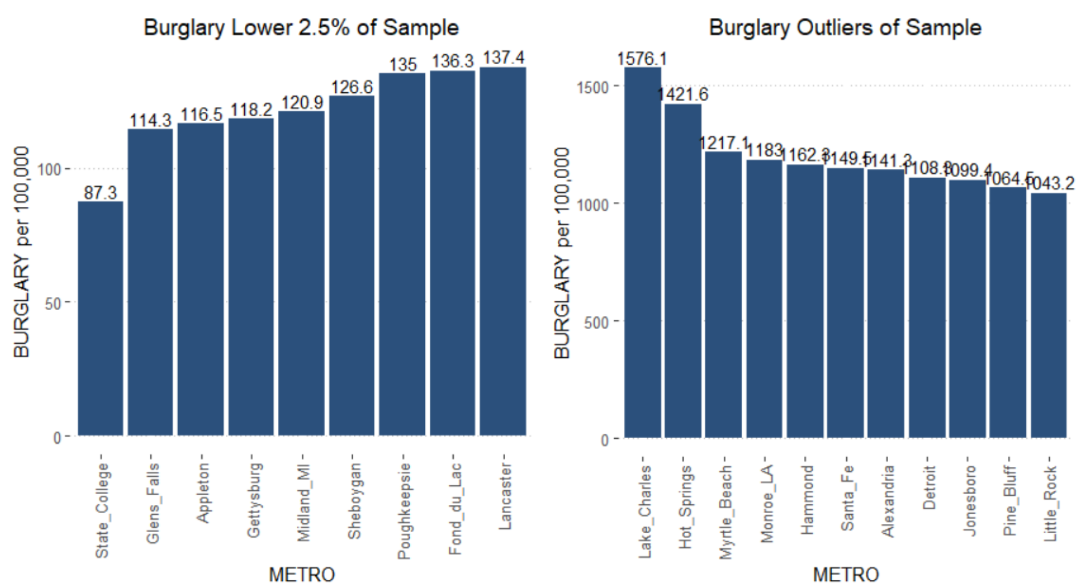


Figure 6.2.10: Burglary Outliers and Lower 2.5% of Sample

The right plot in Figure 6.2.10 contains two areas with larger burglary rates per 100,000: Lake Charles (1576.1) and Hot Springs (1421.6). What is noteworthy is these areas have not shown up on any other of the other outlier plots.

6.2.2.6 Larceny Distribution

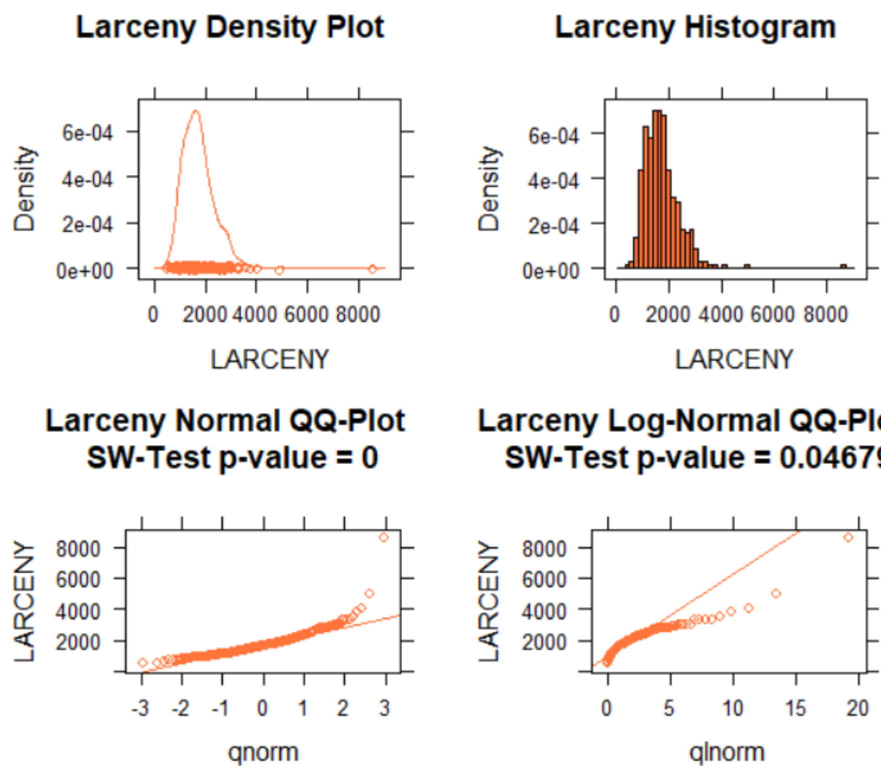


Figure 6.2.11: Larceny Distribution Plots

Based on the density and histogram in Figure 6.2.11, the distribution of Larceny looks right skewed with several outliers. Next, one uses Shapiro-Wilk test for normality and lognormality with $\alpha = 0.1$.

$$H_0 : \text{ Population Distribution for Larceny is Normal}$$

$$H_1 : \text{ Population Distribution for Larceny is not Normal}$$

$$\mathrm{W} = 0.90035; \mathrm{p} - \text{value} \cong 0$$

$$H_0 : \text{ Population Distribution for Larceny is Lognormal}$$

$$H_1 : \text{ Population Distribution for Larceny is not Lognormal}$$

$$\text{p} - \text{value} \cong 0.04679$$

One rejects H_0 for normality and lognormality.

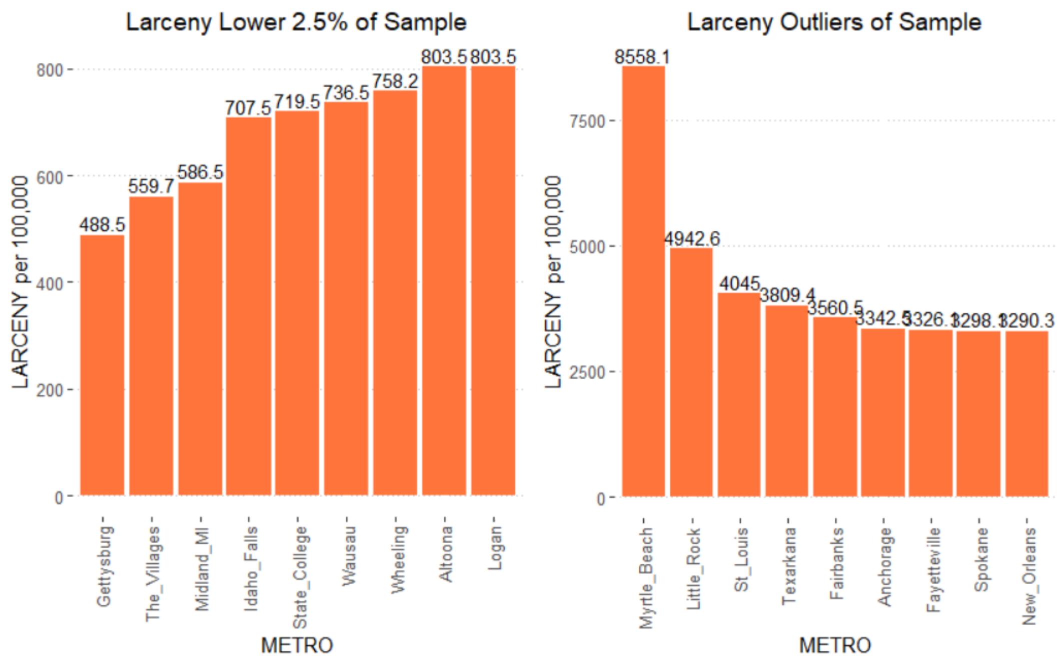


Figure 6.2.12: Larceny Outliers and Lower 2.5% of Sample

The right plot in Figure 6.2.12 has one extreme crime area that stands out compared to the other outliers. Myrtle Beach's (8558.1) larceny crime rate is almost double any other of the outliers. Theft of person property is often higher in tourist destinations. It is surprising that Las Vegas is not one of the high-raking areas for this type of crime.

6.2.2.7 Vehicle Distribution

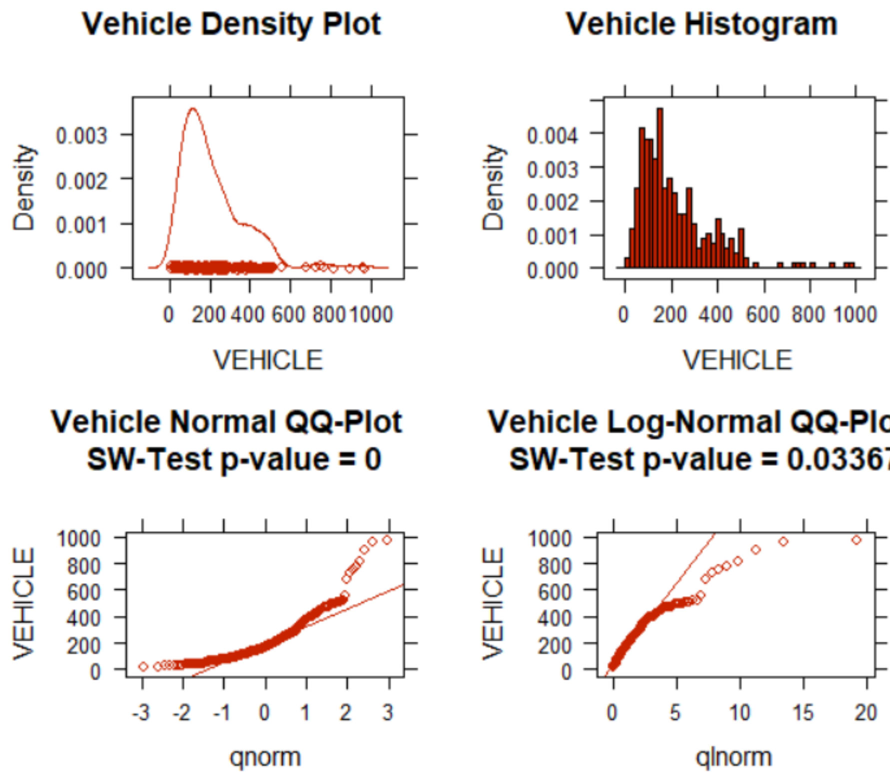


Figure 6.2.13: Vehicle Distribution Plots

Based on the density and histogram in Figure 6.2.11, the distribution of Vehicle looks right skewed with several outliers. Next, one uses Shapiro-Wilk test for normality and lognormality with $\alpha = 0.1$.

$$H_0 : \text{ Population Distribution for Vehicle is Normal}$$

$$H_1 : \text{ Population Distribution for Vehicle is not Normal}$$

$$\mathrm{W} = 0.8565; \mathrm{p} - \mathrm{value} \cong 0$$

$$H_0: \text{ Population Distribution for Vehicle is Lognormal}$$

$$H_1: \text{ Population Distribution for Vehicle is not LogNormal}$$

$$\text{p} - \text{value} \cong 0.3367$$

One rejects H_0 for normality and fails to reject H_0 for lognormality.

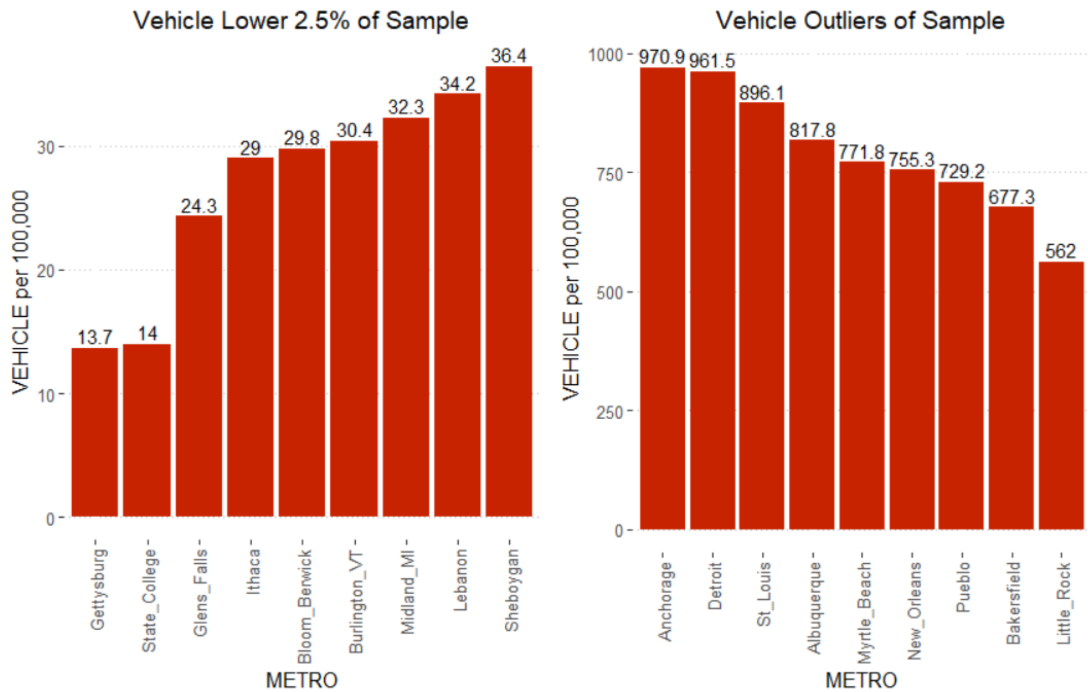


Figure 6.2.14: Vehicle Outliers and Lower 2.5% of Sample

The right plot in Figure 6.2.14 does not show any metropolitan statistical areas where vehicle theft stands out significantly more than others.

6.3 Bivariate Distribution Analysis

6.3.1 Correlation Matrix for US Crime 2018

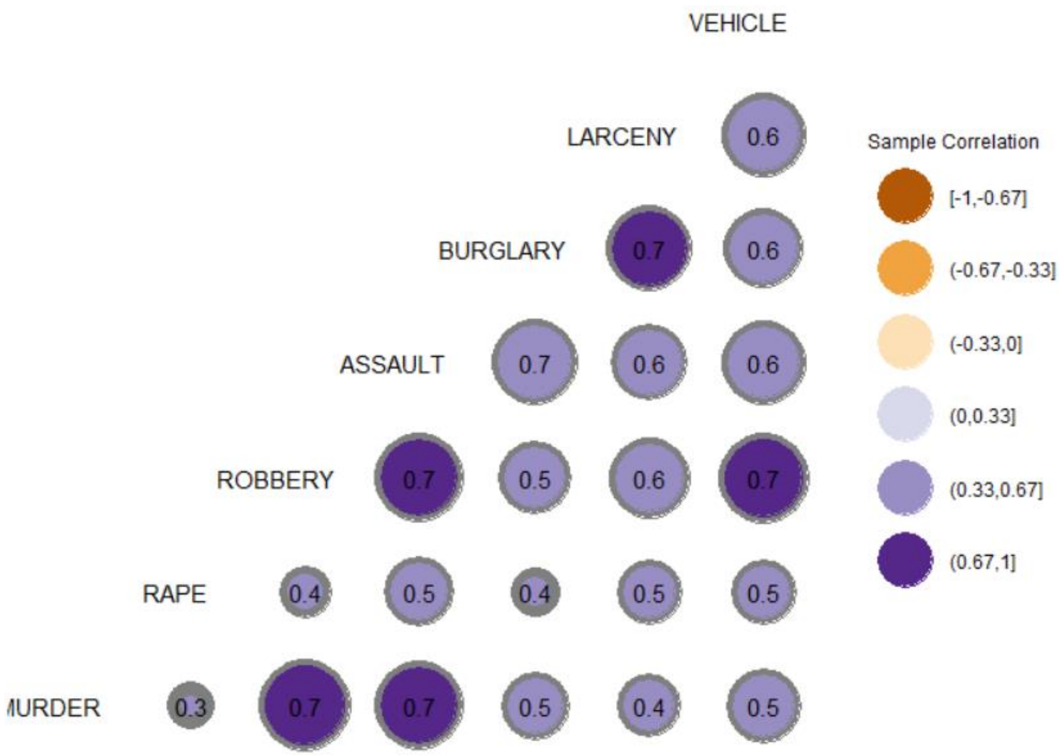


Figure 6.3.1: Correlation Matrix for US Crime 2018

In Figure 6.3.1, one can see strong positive sample correlation between murder and robbery, murder and assault, robbery and assault, burglary and larceny, and robbery and vehicular theft.

6.3.2 Contour-Scatter Matrix

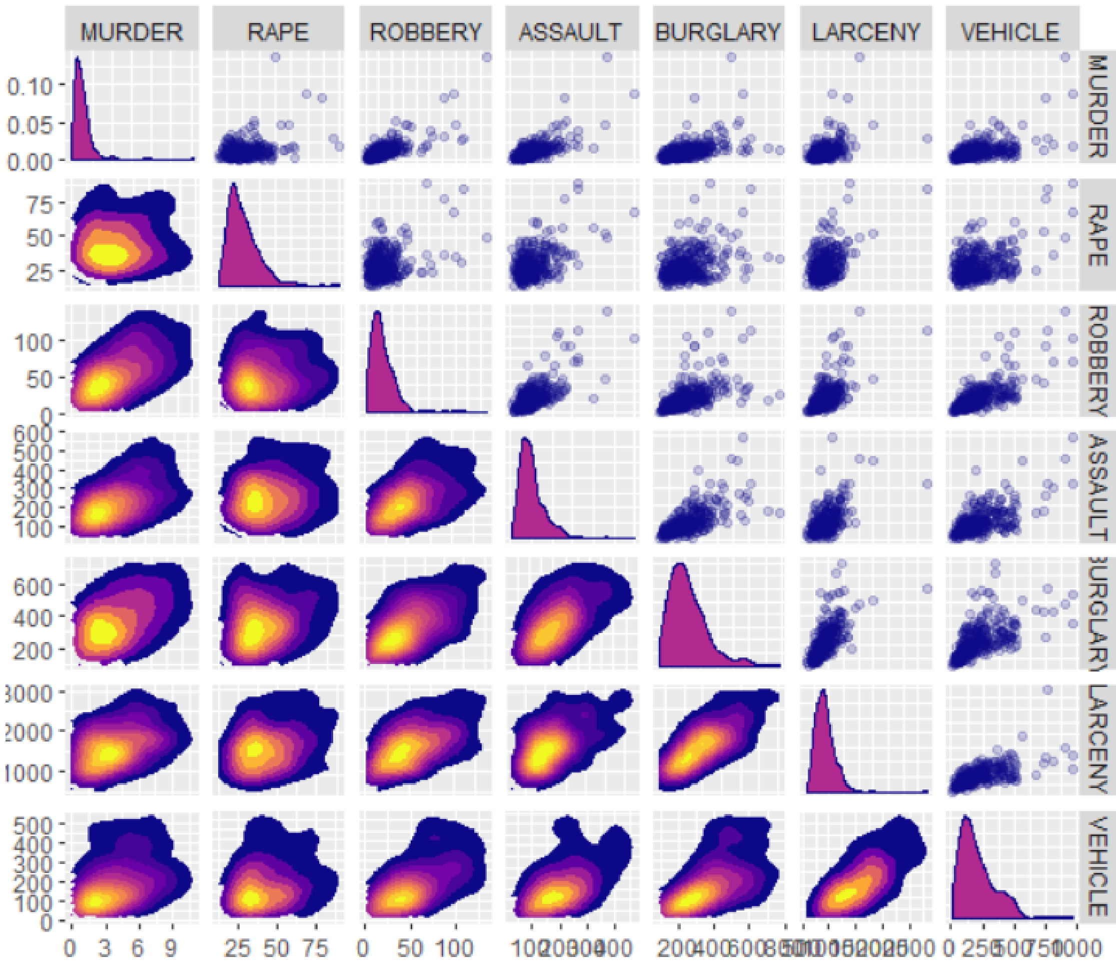


Figure 6.3.2: Contour-Scatter Matrix

The upper diagonal of Figure 6.3.2 displays scatterplots for the seven US Crime 2018 characteristics (variables). One can see there is a dense cloud on the lower-left part of most of the scatterplots linked to areas where pairs of characteristics have lower or medium crime rates. In contrast, one can see less dense scatter in the upper-right of the scatterplots related to those areas where high to extremely crime rates exist.

The lower diagonal of Figure 6.3.2 displays contour plots where the 2-d density is colored with a lighter color for more dense regions and the 2-d density is colored darker for less dense regions. Specifically, the contour plots are a nice way to visualize the bivariate densities in two dimensions instead of in three dimensions. Here, with the contour plots, one can see the densest regions for each pair of variables, unlike in the upper diagonal where it is obscured by the larger number of dots scattered in close proximity.

6.4 Multivariate Distribution Analysis

6.4.1 Testing Multivariate Normality

Using the generalization of Shapiro-Wilk test (Villasenor-Alva and Gonzalez-Estrada 2009) for multivariate normality one can test

$$H_0 : \text{ Population Distribution is Multivariate Normal}$$

$$H_1 : \text{ Population Distribution is not Multivariate Normal}$$

$$\mathrm{W} = 0.8513; \mathrm{p} - \mathrm{value} \cong 0$$

Consequently, one rejects H_0. There is sufficient evidence to say that the US Crime population distribution is not multivariate normal.

6.5 Sample PCA for Standardized US Crime 2018

When individual sample characteristics have vastly different ranges they are routinely standardized before running a principal components analysis [3, p. 439]. Otherwise the characteristics with the largest ranges will dominate the first few sample principal components. Hence, the first step in the principal components analysis is to standardize the US Crime 2018 data.

6.5.1 Descriptives for Standardized US Crime 2018

Table 6.5.1: Descriptives for Standardized US Crime 2018

	vars	n	sd	min	q1	median	mean	q3	max	range
MURDER	1	327	1	-0.94	-0.58	-0.22	0	0.21	10.24	11.18
RAPE	2	327	1	-1.44	-0.68	-0.23	0	0.43	5.66	7.1
ROBBERY	3	327	1	-1.12	-0.60	-0.24	0	0.28	6.5	7.62
ASSAULT	4	327	1	-1.32	-0.65	-0.2	0	0.29	6.62	7.94
BURGLARY	5	327	1	-1.49	-0.74	-0.18	0	0.52	4.89	6.38
LARCENY	6	327	1	-1.72	-0.64	-0.12	0	0.40	9.28	10.99
VEHICLE	7	327	1	-1.25	-0.70	-0.3	0	0.41	4.71	5.96

In Table 6.5.1, one can see that all sample means are 0 and all sample standard deviations are 1. Further, the respective ranges are comparable in size. Now, elements in the standardized multivariate random sample matrix that are positive will be above the sample mean and elements that are negative will be below the sample mean.

6.5.2 Sample PCA for Standardized US Crime 2018

6.5.2.1 Explained Standardized Sample Variance by Principal Component for US Crime 2018

Table 6.5.2: Explained Standardized Sample Variance by Principal Component

	y1	**y2**	**y3**	**y4**	**y5**	**y6**	**y7**
Eigenvalues	4.4138	0.7695	0.6568	0.4503	0.3226	0.2163	0.1707
% of Variance	63.05%	10.99%	9.38%	6.43%	4.61%	3.09%	2.44%
Cumulative %	63.05%	74.05%	83.43%	89.86%	94.47%	97.56%	100.00%

The first row of Table 6.5.2 displays the standardized sample variances for each of the sample principal components ($\text{var}(\hat{Y}_i) = \hat{\lambda}_i$ for $i = 1, \ldots, 7$). The second row provides the percent of standardized sample variance due to the ith sample principal component $\left(\frac{\hat{\lambda}_i}{7} \cdot 100\%, i = 1, \ldots, 7\right)$. Finally, the third row shows the percent of standardized sample variance due to the first kth sample principal component $\left(\frac{\sum_{i=1}^{k} \hat{\lambda}_i}{7},\ k \leq 7\right)$. One can see that the first three sample principal components account for 83.43% of the total standardized variation in the sample from US Crime 2018. Figure 6.5.1 gives us a way to visualize the relation between the standardized sample principal components and their percentages of explained standardized sample variance.

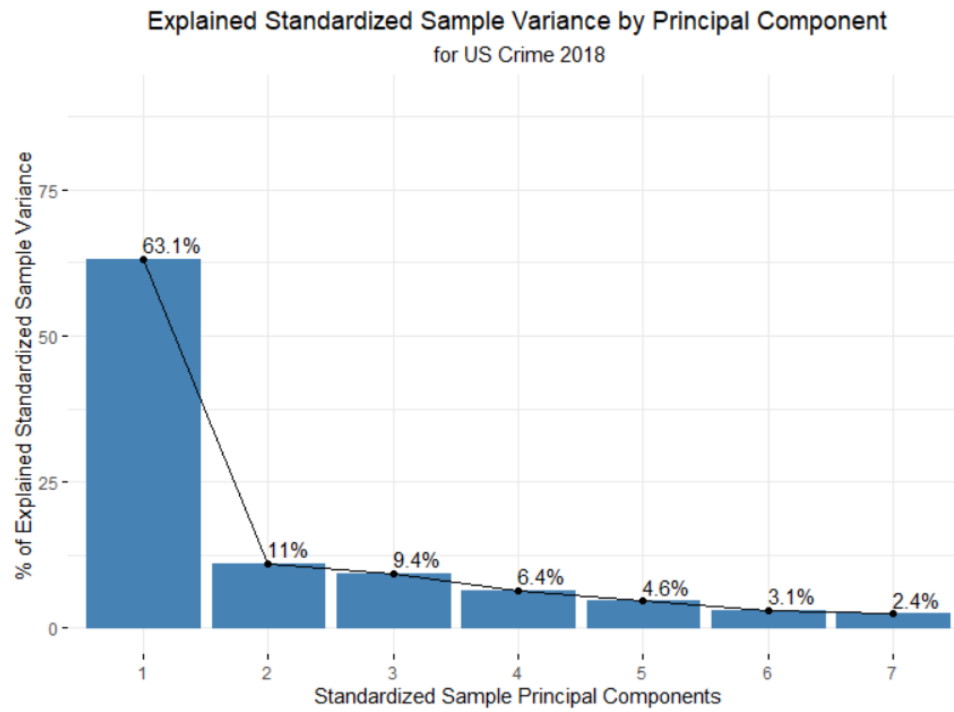


Figure 6.5.1: Explained Standardized Sample Variance by Principal Component

6.5.2.2 Sample Principal Components for Standardized Crime 2018

Table 6.5.3: Sample Principal Components for Standardized Crime 2018

	y1	y2	y3	y4	y5	y6	y7
MURDER	-0.3704	0.4863	-0.3767	0.197	-0.3164	-0.4799	0.3394
RAPE	-0.2849	-0.7469	-0.5518	0.0673	-0.0436	-0.1588	-0.1574
ROBBERY	-0.4087	0.3387	-0.131	-0.3787	-0.1169	0.1769	-0.7161
ASSAULT	-0.4144	0.0985	-0.1529	0.4294	0.2454	0.709	0.2191
BURGLARY	-0.372	-0.0847	0.555	0.5261	0.1749	-0.3644	-0.3261
LARCENY	-0.3792	-0.2702	0.4419	-0.2523	-0.6337	0.193	0.2923
VEHICLE	-0.4009	-0.0313	0.0989	-0.5371	0.6261	-0.2013	0.328

Given that the first three sample principal components yield 83.43% of the total standardized variation in the sample, there is no need to use the other four sample components in one's analysis.

attempted to be explained in the context of the subject matter. To demonstrate, sample principal component $\hat{y}_1$ has eigenvector components of roughly equal magnitudes. Thus, $\hat{y}_1$ can be considered a *general crime component*. If one was to explain a metropolitan statistical area's crime rate with one value, then $\hat{y}_{j1}$ would be it. Most importantly because $\hat{y}_1$ maximizes the standardized sample variance $\text{var}(\hat{y}_1)$ subject to $\underset{(1\times 7)}{\hat{\mathbf{e}}_1'} \cdot \underset{(7\times 1)}{\hat{\mathbf{e}}_1} = 1$ and $\text{cov}(\hat{y}_1, \hat{y}_k) = 0, k = 2, \ldots, 7$. Notice that all the eigenvector components are negative; accordingly, an area with larger crime rates would have a very negative value (in general).

$$\begin{aligned}\hat{y}_1 = -0.37x_{\text{MURDER}} - 0.28x_{\text{RAPE}} - 0.41x_{\text{ROBBERY}} - 0.41x_{\text{ASSAULT}} \\ -0.37x_{\text{BURGLARY}} - 0.37x_{\text{LARCENY}} - 0.40x_{\text{VEHICLE}}\end{aligned}$$

with jth observation

$$\begin{aligned}\hat{y}_{j1} = -0.37x_{j,\text{MURDER}} - 0.28x_{j,\text{RAPE}} - 0.41x_{j,\text{ROBBERY}} - 0.41x_{j,\text{ASSAULT}} \\ -0.37x_{j,\text{BURGLARY}} - 0.37x_{j,\text{LARCENY}} - 0.40x_{j,\text{VEHICLE}}\end{aligned}$$

Sample principal component $\hat{y}_2$ has largest eigenvector component magnitudes on murder and rape. Therefore, $\hat{y}_2$ could be deemed a *heinous crime component*. If the area has a much larger murder rate, then rape rate, $\hat{y}_{j2}$ will likely stand out in the positive direction. If the area has a much larger rape rate, then murder rate, $\hat{y}_{j2}$ will likely stand out in the negative direction. If the area has approximately equal values, then $\hat{y}_{j2}$ will likely not stand out in either direction.

$$\hat{y}_2 = 0.49x_{\text{MURDER}} - 0.75x_{\text{RAPE}} + 0.34x_{\text{ROBBERY}} + 0.10x_{\text{ASSAULT}}$$

$$-0.08x_{\text{BURGLARY}} - 0.27x_{\text{LARCENY}} - 0.03x_{\text{VEHICLE}}$$

with jth observation

$$\hat{y}_{j2} = 0.49x_{j,\text{MURDER}} - 0.75x_{j,\text{RAPE}} + 0.34x_{j,\text{ROBBERY}} + 0.10x_{j,\text{ASSAULT}}$$

$$-0.08x_{j,\text{BURGLARY}} - 0.27x_{j,\text{LARCENY}} - 0.03x_{j,\text{VEHICLE}}$$

Sample principal component $\hat{y}_3$ has negative eigenvector components for violent crime and positive eigenvector components for property crime. Immediately, $\hat{y}_3$ can be thought of as a *crime type component*. That is, areas with particularly negative $\hat{y}_{j3}$ values will often have larger violent crime relative to property crime. Conversely, areas with larger property crime relative to violent crime will have more positive $\hat{y}_{j3}$ values.

$$\hat{y}_3 = -0.38x_{\text{MURDER}} - 0.55x_{\text{RAPE}} - 0.13x_{\text{ROBBERY}} - 0.15x_{\text{ASSAULT}}$$

$$+0.56x_{\text{BURGLARY}} + 0.44x_{\text{LARCENY}} + 0.10x_{\text{VEHICLE}}$$

with jth observation

$$\hat{y}_{j3} = -0.38x_{j,\text{MURDER}} - 0.55x_{j,\text{RAPE}} - 0.13x_{j,\text{ROBBERY}} - 0.15x_{j,\text{ASSAULT}}$$

$$+0.56x_{j,\text{BURGLARY}} + 0.44x_{j,\text{LARCENY}} + 0.10x_{j,\text{VEHICLE}}$$

Note that explaining these principal components is not a perfect science and caution should be exercised when interpreting the $\hat{y}_i$'s in context of the data. Figure 6.5.2 gives a graphical interpretation of how the standardized characteristics contributed to the first three sample principal component derived from the US Crime 2018 data.

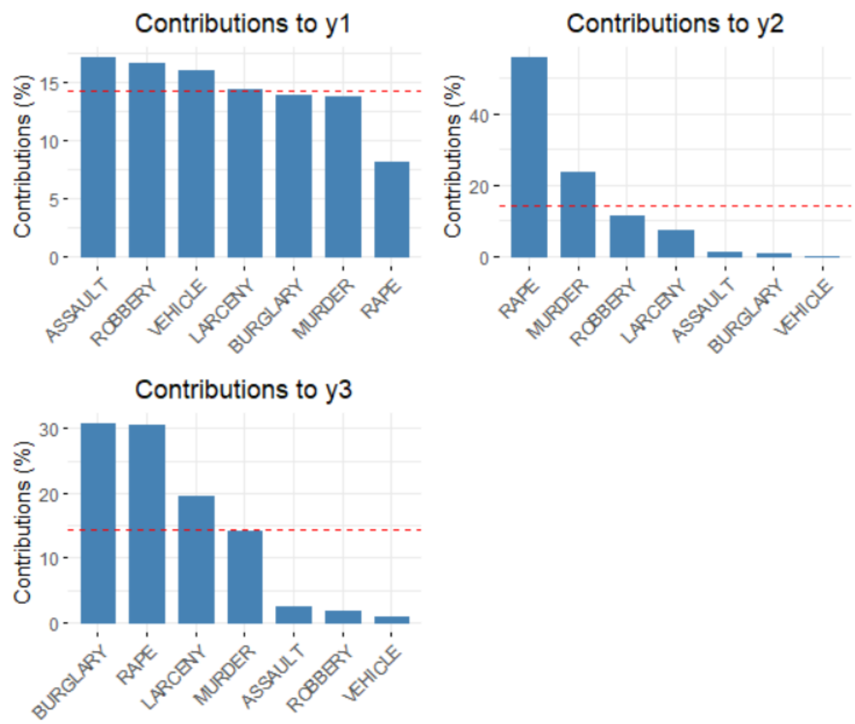


Figure 6.5.2: Contrib. of Standardized Characteristics to Each Principal Component

In Figure 6.5.2, the percent contribution of the kth standardized characteristic to the ith sample principal component is calculated as

$$\text{Sample Contribution}_{ki} = \hat{e}_{ki}^2 \cdot 100\%$$

for $k, i = 1,2, \dots, p$ because $\hat{\mathbf{e}}_i' \cdot \hat{\mathbf{e}}_i \cdot 100\% = 1 \cdot 100\% = 100\%$. Hence, $\hat{e}_{ki}^2$ is the proportion contribution of the kth standardized characteristic to the ith sample principal component. To clarify further, $\hat{\mathbf{e}}_i' \cdot \hat{\mathbf{e}}_i$ represents the squared length or magnitude of the vector $\hat{\mathbf{e}}_i$ so $\hat{e}_{ki}^2 = \hat{e}_{ki} \cdot \hat{e}_{ki}$ is the part that the standardized characteristic z_k that contributes to magnitude of, or squared length of, $\hat{\mathbf{e}}_i$.

6.5.2.3 Correlation Matrix for Sample Principal Components and Standardized Crime 2018 Characteristics

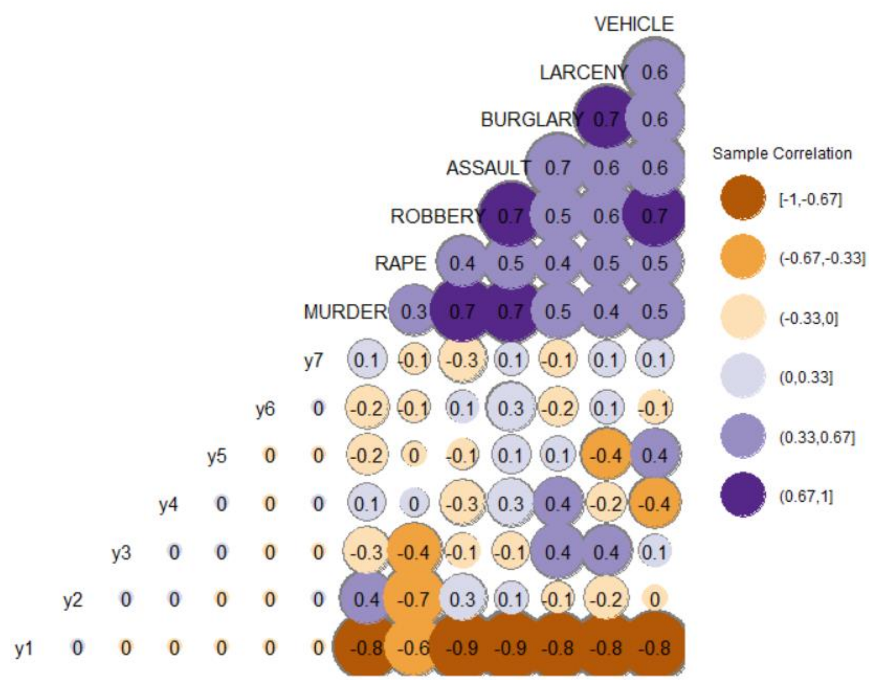


Figure 6.5.3: Correlation Matrix for Sample Principal Components and Standardized Crime 2018 Characteristics

The upper-right triangle of Figure 6.5.3 displays the sample correlations between the (Standardized) US Crime 2018 characteristics as seen in Figure 6.3.1.The bottom-left triangle of Figure 6.5.3 shows the sample principal components are indeed uncorrelated because $\text{cov}(\hat{y}_i, \hat{y}_k) = 0 \; \forall \; i \neq k$.

In the right-bottom square of Figure 6.5.3, the correlations between the sample principal components and the standardized US Crime 2018 characteristics, can be seen. The interpretation of these sample correlations can lead to similar interpretations as looking at $\hat{e}_{ki}$ directly, but with some data, this is not true [3, p. 434]. For Figure 6.5.3, the correlations between sample principal components and the standardized US Crime 2018 characteristics match the original interpretations of the $\hat{e}_{ki}$'s.

To illustrate, the eigenvector components of $\hat{y}_1$ are all negative and nearly the same magnitude. Analogously, the correlations between the eigenvector components of $\hat{y}_1$ and the standardized US Crime 2018 characteristics are all strong negatively correlated. For the eigenvector components of $\hat{y}_2$ and the standardized US Crime 2018 characteristics, one can see a strong negative correlation between the standardized rape characteristic and its respective eigenvector component. In the same way, the standardized murder characteristic is positively correlated with its eigenvector counterpart. Principal component $\hat{y}_3$ has negative correlations with the violent crime characteristics and positive correlations with the property crime characteristics. Henceforth, the correlation structure between the sample principal components and the standardized US Crime 2018 characteristics agree with the signs and magnitudes of the $\hat{e}_{ki}$'s.

6.5.2.4 Scatterplots for Sample Principal Components from Standardized US Crime 2018

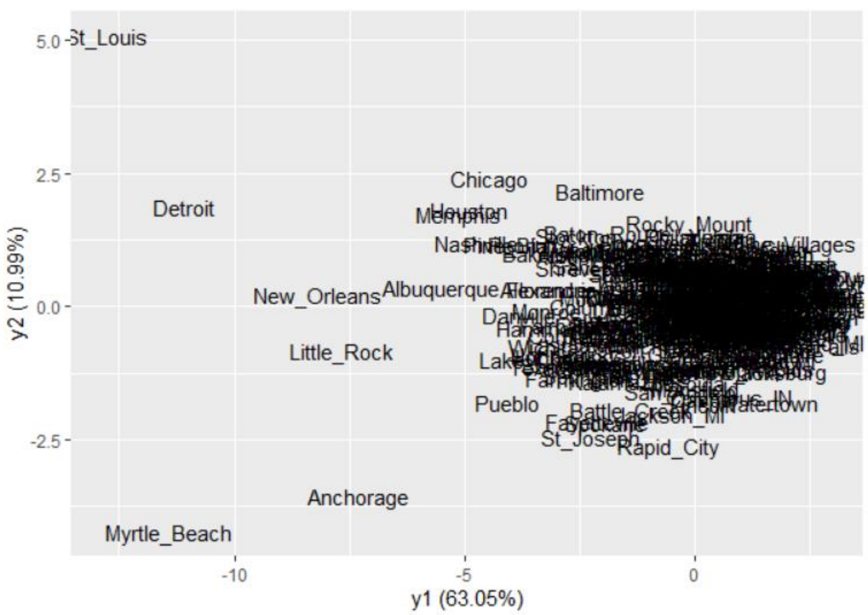


Figure 6.5.4: Scatterplot for $\hat{y}_2 \sim \hat{y}_1$

Figure 6.5.4 plots sample principal components

$$\hat{y}_{j2} = 0.49x_{j,\text{MURDER}} - 0.75x_{j,\text{RAPE}} + 0.34x_{j,\text{ROBBERY}} + 0.10x_{j,\text{ASSAULT}}$$
$$-0.08x_{j,\text{BURGLARY}} - 0.27x_{j,\text{LARCENY}} - 0.03x_{j,\text{VEHICLE}}$$

by

$$\hat{y}_{j1} = -0.37x_{j,\text{MURDER}} - 0.28x_{j,\text{RAPE}} - 0.41x_{j,\text{ROBBERY}} - 0.41x_{j,\text{ASSAULT}}$$
$$-0.37x_{j,\text{BURGLARY}} - 0.37x_{j,\text{LARCENY}} - 0.40x_{j,\text{VEHICLE}}$$

for $j = 1, 2, \ldots, 327$.

From Figure 6.5.4, metropolitan statistical areas to the far left in the $\hat{y}_1$ direction are those places with very extreme crimes rates on one or more characteristics. Specifically, because $\hat{y}_1$ has all negative eigenvector components, areas with large crime rates will have sample principle components scores far to the left. Thus, St. Louis, Detroit, New Orleans, Little Rock, Anchorage, and Myrtle Beach can be put into the severe crime category based on the *general crime component* $\hat{y}_1$.

Next, from Figure 6.5.4, metropolitan statistical areas in the upper region of $\hat{y}_2$ dimension are going to have high murder rates relative to rape rates. These areas include St. Louis, Chicago, and Baltimore (see also Figure 6.6.2 for Murder Outliers). At the same time, metropolitan statistical areas in the lower region of $\hat{y}_2$ are going to have high rape rates relative to murder rates. These areas include Myrtle Beach and Anchorage (see also Figure 6.6.3 for Rape Outliers). After all, $\hat{y}_2$ is the *heinous crime component*, which is dominated by the negative eigenvector component for rape and the positive eigenvector component for murder.

There are also cases where areas had large murder and rape rates that ended up in the center region of $\hat{y}_2$. These areas include Detroit, New Orleans, and Little Rock (see Figure 6.6.2-6.6.3). Finally, areas that had smaller crimes rates would end up center around $(\hat{y}_1 = 0, \hat{y}_2 = 0)$.

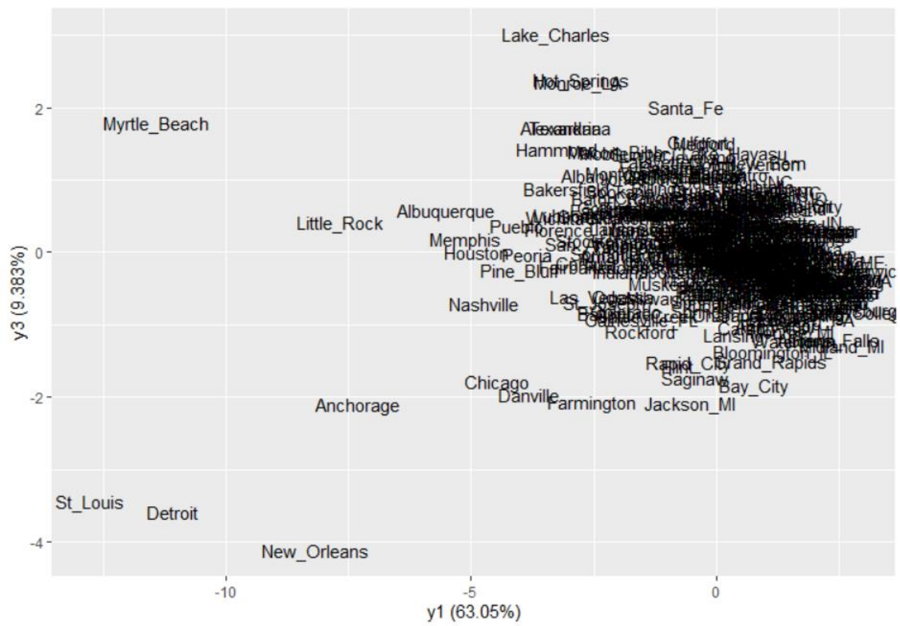


Figure 6.5.5: Scatterplot for $\hat{y}_3 \sim \hat{y}_1$

Figure 6.5.5 plots sample principal components

$$\hat{y}_{j3} = -0.38x_{j,\text{MURDER}} - 0.55x_{j,\text{RAPE}} - 0.13x_{j,\text{ROBBERY}} - 0.15x_{j,\text{ASSAULT}}$$
$$+0.56x_{j,\text{BURGLARY}} + 0.44x_{j,\text{LARCENY}} + 0.10x_{j,\text{VEHICLE}}$$

by

$$\hat{y}_{j1} = -0.37x_{j,\text{MURDER}} - 0.28x_{j,\text{RAPE}} - 0.41x_{j,\text{ROBBERY}} - 0.41x_{j,\text{ASSAULT}}$$
$$-0.37x_{j,\text{BURGLARY}} - 0.37x_{j,\text{LARCENY}} - 0.40x_{j,\text{VEHICLE}}$$

for $j = 1, 2, \dots, 327$.

From Figure 6.5.5, metropolitan statistical areas in the upper region of $\hat{y}_3$ have serious crime rates related to one or more violent crimes relative to property crimes. One the other hand, metropolitan statistical areas in the lower region of $\hat{y}_3$ have significant crime rates related to one or more property crimes relative to violent crimes. That is, $\hat{y}_3$ has negative eigenvector components for violent crime and positive eigenvector components for property crime. Specifically, for violent crime $\hat{y}_3$ is most weighted towards murder and rape. While, property crime is most weighted towards burglary and larceny. This is the *crime type component*.

Lake Charles has the largest value on $\hat{y}_3$. It is interesting because the area only came up once in the outliers for burglary where it had the largest number of burglaries (1576.1) per 100, 000 in the nation (see Figure 6.2.10). Otherwise, Lake Charles has not shown up on one's radar.

Myrtle Beach is interesting because it has large crime rates for all characteristics except for murder. Thus, it is tough to say whether Myrtle Beach is worse with respect to violent crime or property crime based on its $\hat{y}_{j3}$ value. In short, $\hat{y}_3$ has neutralized the effect for Myrtle Beach.

St. Louis, Detroit, and New Orleans have high crime rates on most of the characteristics, but violent crime is most pronounced in $\hat{y}_3$. Most notably, St. Louis has the largest murder rate of 60.9, Detroit has the second highest murder rate at 38.9, and New Orleans has the third highest murder rate at 37.1. New Orleans ranks third in rape at 171.8 and Detroit ranks fourth at 147.2. St Louis leads in robbery with 473.2, Detroit takes fourth with 344, and New Orleans in sixth with 307.5. Detroit is in first for assault with 1477.8 and St. Louis is in second with 1165.6.

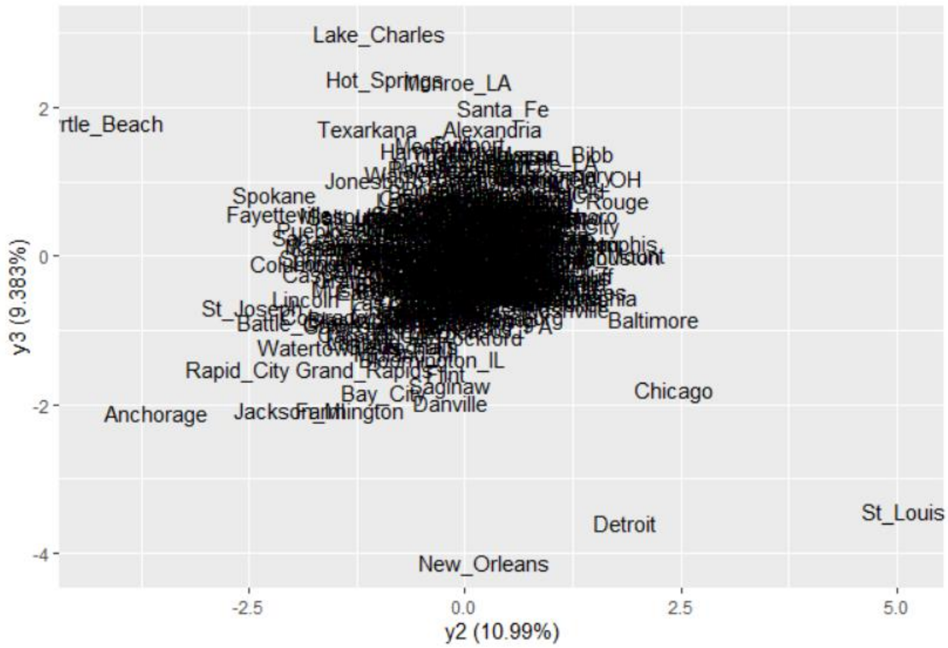


Figure 6.5.6: Scatterplot for $\hat{y}_3 \sim \hat{y}_2$

Figure 6.5.6 plots sample principal components

$$\hat{y}_{j3} = -0.38x_{j,\text{MURDER}} - 0.55x_{j,\text{RAPE}} - 0.13x_{j,\text{ROBBERY}} - 0.15x_{j,\text{ASSAULT}}$$
$$+0.56x_{j,\text{BURGLARY}} + 0.44x_{j,\text{LARCENY}} + 0.10x_{j,\text{VEHICLE}}$$

by

$$\hat{y}_{j2} = 0.49x_{j,\text{MURDER}} - 0.75x_{j,\text{RAPE}} + 0.34x_{j,\text{ROBBERY}} + 0.10x_{j,\text{ASSAULT}}$$
$$-0.08x_{j,\text{BURGLARY}} - 0.27x_{j,\text{LARCENY}} - 0.03x_{j,\text{VEHICLE}}$$

for $j = 1, 2, \dots, 327$.

6.6 k-Means Clustering Method

The k-Means clustering algorithm is used to partition a set of n unclassified multivariate sample observations $\underset{(p\times 1)}{\mathbf{x}_1}, \underset{(p\times 1)}{\mathbf{x}_2}, \dots, \underset{(p\times 1)}{\mathbf{x}_n}$ into k clusters or groups using a distance metric, most commonly, Euclidean distance. Note that the number of clusters k must be specified in advance, which there are various numerical processes to help, analytically, specify this parameter [10, p. 532].

Because the k-Means algorithm, by default, uses the Euclidean distance metric it suffers from certain deficiencies based on the number of calculations it must make and the size of those calculations. Respectively, the k-Means algorithm runs slower and has trouble finding reasonable clusters in the same proximity when:

(1) n and p are large.

(2) The ranges of the $x_1, x_2, \dots, x_p$ are large and/or when the ranges of $x_1, x_2, \dots, x_p$ are largely different from each other.

One solution to solve the range dilemma is to standardize the sample and use $\underset{(p\times 1)}{\mathbf{z}_1}, \underset{(p\times 1)}{\mathbf{z}_2}, \dots, \underset{(p\times 1)}{\mathbf{z}_n}$ as the inputs into the k-Means algorithm. However, this solution does not address the number of characteristics p being large. To address this issue, one can subset p variables in some meaningful way and continue with the k-Means analysis; but it is in generally difficult to make the decision of which characteristics to keep and which to lose. However, another option exists to solve both problems simultaneously. Specifically, one can use the first two or three sample principal components from the standardized sample provided that they account for a large proportion of the variability in $\underset{(p\times 1)}{\mathbf{z}_1}, \underset{(p\times 1)}{\mathbf{z}_2}, \dots, \underset{(p\times 1)}{\mathbf{z}_n}$.

For the US Crime 2018 data, we will use the standardized sample and the first three sample principal components derived from the standardized sample as inputs into the k-Means algorithm to compare. One can then see how similar or different the two inputs behave with respect to the k-Means cluster assignments.

6.6.1 Choosing k

One black-box method for choosing the appropriate k for several clustering methods is found in the R package NbClust. NbClust provides 30 indices for determining the relevant number of clusters and proposes to users the best clustering scheme from the different results obtained by varying all combinations of number of clusters, distance measures, and clustering methods. It can simultaneously compute all the indices and determine the number of clusters in a single function call [11].

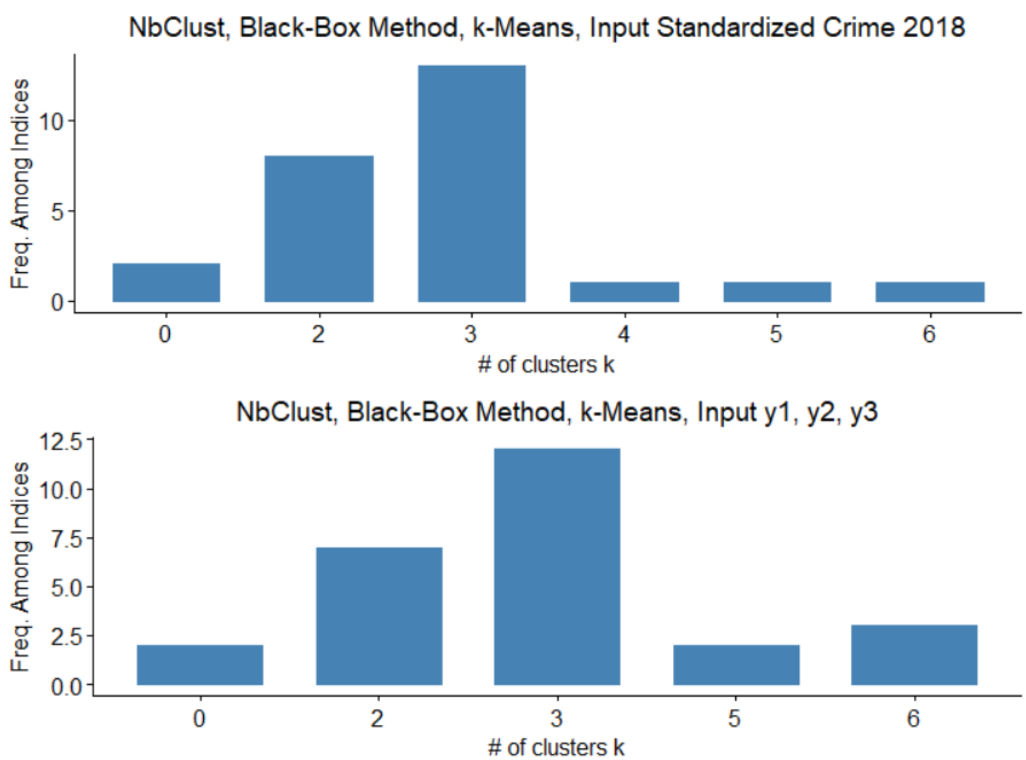


Figure 6.6.1: NbClust, Black-Box Method, k-Means

In Table 6.6.1, the optimal number of clusters is found to be $k = 3$ for both inputs, Standardized Crime 2018 and $\hat{y}_1, \hat{y}_2, \hat{y}_3$.

6.6.2 $\boldsymbol{k}$-Means, $\boldsymbol{k = 3}$

6.6.2.1 $\boldsymbol{k}$-Means, $\boldsymbol{k = 3}$, Cluster Sizes

Table 6.6.1: k-Means, $k = 3$, Cluster Sizes

k=3, k-Means, Standardized Crime 2018, Cluster Size			
cluster	**1**	**2**	**3**
size	11	116	200
k=3, k-Means, y1, y2, y3, Cluster Size			
cluster	**1**	**2**	**3**
size	6	109	212

6.6.2.2 $\boldsymbol{k}$-Means, $\boldsymbol{k = 3}$, Differences in Cluster Assignments

Table 6.6.2: k-Means, $k = 3$, Differences in Cluster Assignments

<table>
<tr><th colspan="5">k-Means, k=3, Differences in Cluster Assignments</th></tr>
<tr><td></td><td colspan="4">y1, y2, y3</td></tr>
<tr><td rowspan="4">Standardized Crime 2018</td><td>cluster</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td>1</td><td>6</td><td>5</td><td>0</td></tr>
<tr><td>2</td><td>0</td><td>104</td><td>12</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>200</td></tr>
</table>

From Table 6.6.2, one can see that 5 metropolitan statistical areas were assigned to cluster 1 using Standardized Crime 2018 and the same 5 metropolitan statistical areas where assigned to cluster 2 using $\hat{y}_1, \hat{y}_2, \hat{y}_3$. Similarly, one can see that the same 12 metropolitan statistical areas were assigned to cluster 2 using Standardized Crime 2018 and cluster 3 using $\hat{y}_1, \hat{y}_2, \hat{y}_3$. Table 6.6.3 presents the specific metropolitan statistical areas assigned to different clusters. These cases are usually located near the border's edges of the cluster regions.

Table 6.6.3: k-Means, $k = 3$, Differences in Cluster Assignments for Clusters 1,2,3

k-Means, k=3, Differences in Assignments for Cluster 1 and 2		
Metropolitan Statistical Area	**Standardized Crime 2018**	**y1, y2, y3**
Albuquerque	1	2
Chicago	1	2
Houston	1	2
Memphis	1	2
Nashville	1	2
k-Means, k=3, Differences in Assignments for Cluster 2 and 3		
Metropolitan Statistical Area	**Standardized Crime 2018**	**y1, y2, y3**
Brunswick	2	3
Charleston	2	3
Columbus_OH	2	3
Dayton	2	3
Jackson_MI	2	3
Lexington	2	3
Orlando	2	3
Reno	2	3
Saginaw	2	3
Salem	2	3
San_Jose	2	3
Honolulu	2	3

6.6.2.3 k-Means, $k = 3$, Sample Cluster Mean Vectors

Table 6.6.4: k-Means, $k = 3$, Sample Cluster Mean Vectors

k-Means, k=3, Standardized Crime 2018, Sample Cluster Mean Vectors							
cluster	**MURDER**	**RAPE**	**ROBBERY**	**ASSAULT**	**BURGLARY**	**LARCENY**	**VEHICLE**
1	22.7091	111.782	312.0364	865.3273	811.3909	3678.564	681.509
2	6.79483	58.5103	97.13879	362.3543	616.0431	2158.441	318.863
3	3.1705	43.2875	41.8045	183.871	310.761	1404.355	129.394
k-Means, k=3, y1, y2, y3, Sample Cluster Mean Vectors							
cluster	**MURDER**	**RAPE**	**ROBBERY**	**ASSAULT**	**BURGLARY**	**LARCENY**	**VEHICLE**
1	29.55	153.2	316.9833	1010	925.7	4402.25	819.6
2	7.30917	58.8835	108.5073	391.3817	640.5229	2227.902	332.594
3	3.29293	44.0415	44.01981	186.8175	316.8269	1426.697	137.703
Original Sample Mean Vector for Crime 2018							
	MURDER	**RAPE**	**ROBBERY**	**ASSAULT**	**BURGLARY**	**LARCENY**	**VEHICLE**
	5.11	50.99	70.52	270.11	435.9	1748.36	215.18

Table 6.6.4 reveals that cluster 3 has smaller sample mean components for both inputs compared to the original sample mean vector for Crime 2018. Thus, cluster 3 can be labeled the *below average crime cluster*. Cluster 2 has larger sample mean components for both inputs compared to the original sample mean vector for Crime 2018. Hence, cluster 2 can be labeled the *above average crime cluster*. Cluster 1 has much larger sample mean components for both inputs compared to their respective cluster 2's, cluster 3's, and the original sample mean vector for Crime 2018. Correspondingly, cluster 1 can be labeled the *extreme crime cluster*. At the same time, one should notice that the sample means with input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ are higher than the samples means with input Standardized Crime 2018. The reason will be evident once we plot the cluster assignments on the $\hat{y}_1, \hat{y}_2, \hat{y}_3$ and the original dimensions.

6.6.2.4 k-Means, $k = 3$, Scatterplots on $\hat{y}_1, \hat{y}_2, \hat{y}_3$

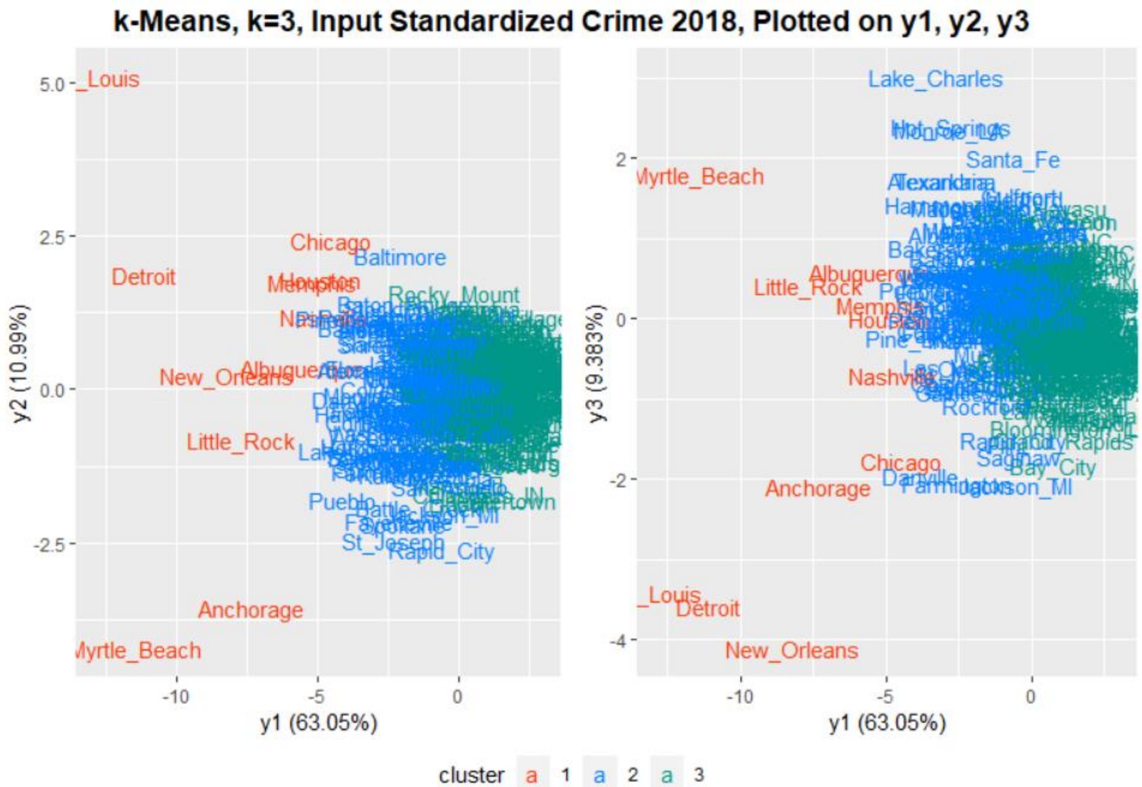


Figure 6.6.1: k-Means, $k = 3$, Input Standardized Crime 2018, Plotted on $\hat{y}_1, \hat{y}_2, \hat{y}_3$

Focusing our attention on the $\hat{y}_1$ (horizontal) dimension or the *general crime component* of Figure 6.6.1, with Standardized Crime 2018 inputs, one can see that the three clusters are fairly well-separated. Cluster 1, the *extreme crime cluster* is farthest to the left because the eigenvector coefficients of $\hat{y}_1$ are negative, making areas with extreme crime on one or more of the characteristics shift to the left. Continuing to focus our attention on $\hat{y}_1$, cluster 2, the *above average crime cluster*, is shifted to the right from clusters 1. We saw in Table 6.6.4, that cluster 2, had smaller sample mean vector components then cluster 1; thus, it makes sense that is would be farther to the right in the $\hat{y}_1$ dimension. Likewise, cluster 3 the *below average crime cluster*, is farther to the right then clusters 1 and 2 given its smaller mean vector components.

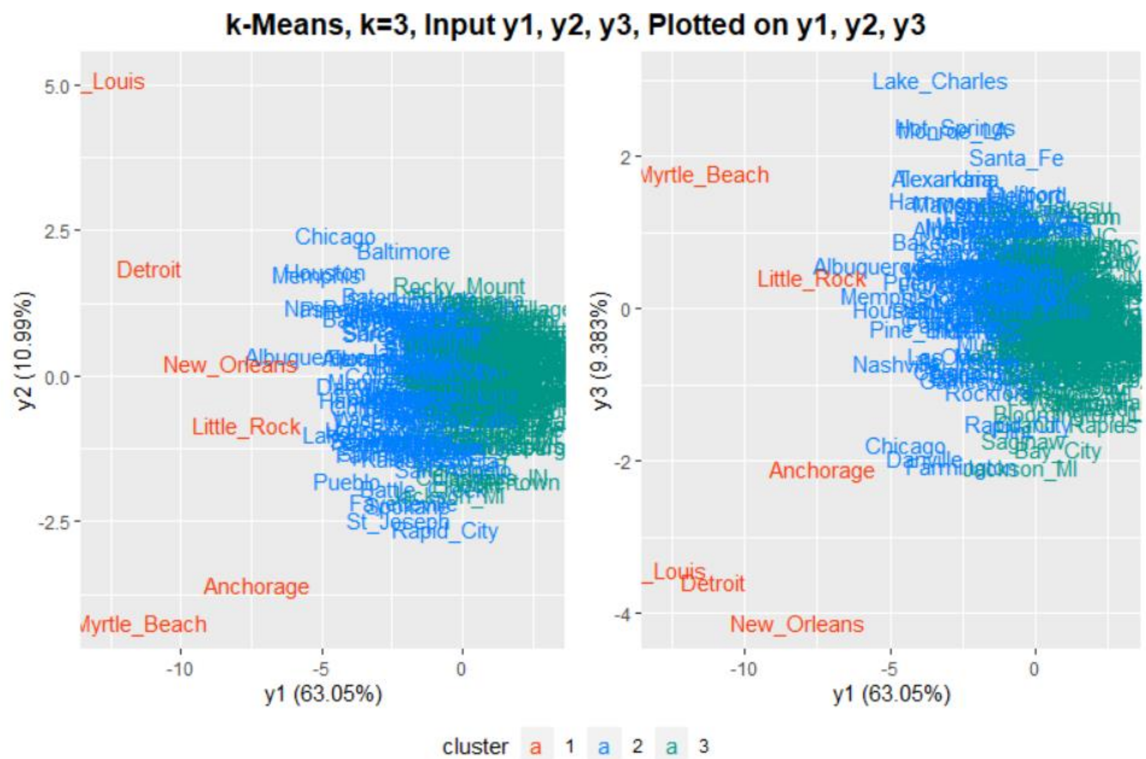


Figure 6.6.2: k-Means, $k = 3$, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, Plotted on $\hat{y}_1, \hat{y}_2, \hat{y}_3$

Referring to Figure 6.6.2, the clusters with input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ do not look remarkably different from clusters in Figure 6.6.1, with input Standardized Crime 2018. Except that cluster 1, the *extreme crime cluster*, has lost five metropolitan statistical areas, Albuquerque, Chicago, Houston, Memphis, and Nashville which have been absorbed into cluster 2 the *above average crime cluster*. These areas have large crime rates but not as extreme as St. Louis, Detroit, New Orleans, Little Rock, Anchorage, and Myrtle Beach with respect to the point estimate $\hat{y}_{j1}$. That is why the cluster mean vector components are larger for the *extreme crime cluster* with input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ compared to the *extreme crime cluster* with input Standardized Crime 2018. Lastly, one should mention that Albuquerque, Chicago, Houston, Memphis, and Nashville are on the boundary of clusters 1 and 2 for both inputs; consequently, being assigned to either cluster does not seem unreasonable.

6.6.2.5 k-Means, $k = 3$, Scatterplots on Original Crime 2018 Dimensions

Another method of visualizing the k-Means, $k = 3$, cluster assignments for inputs Standardized Crime 2018 and $\hat{y}_1, \hat{y}_2, \hat{y}_3$ is to plot them using a scatterplot matrix on the original Crime 2018 dimensions.

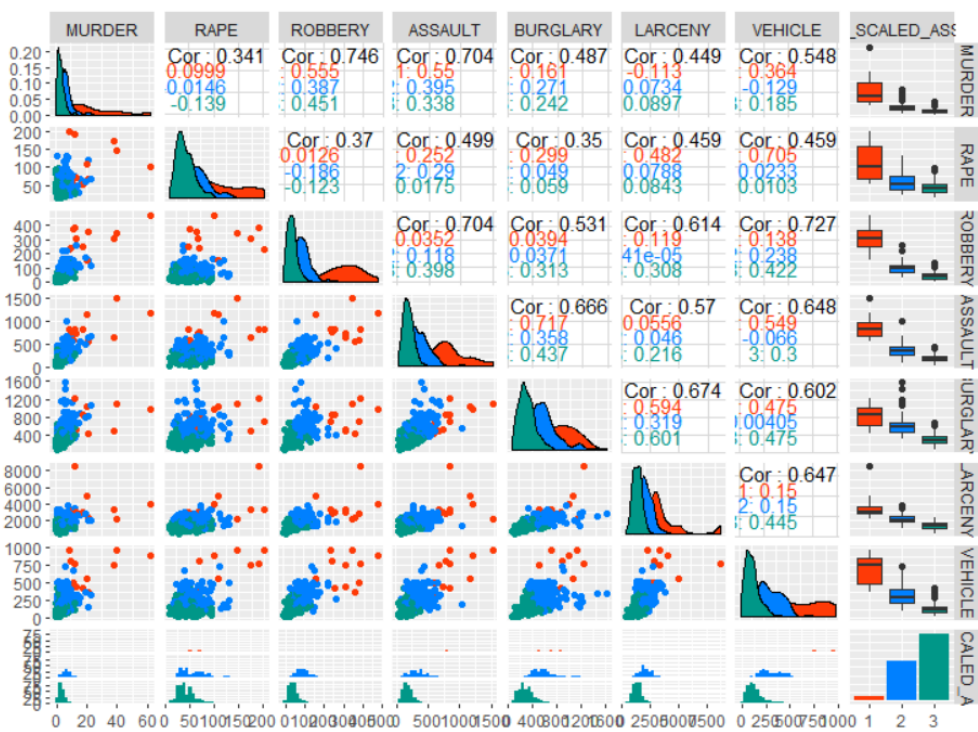


Figure 6.6.3: k-Means, $k = 3$, Input Standardized Crime 2018, Original Crime 2018

Looking at Figure 6.6.3, k-Means, $k = 3$, Input Standardized Crime 2018, one can see the densities for each cluster on each characteristic. Cluster 1's distributions are all shifted farthest to the right giving it the largest sample mean on each characteristic. Next, cluster 2 has the second largest sample means based on the position of the densities. Afterward, cluster 3 has the smallest sample means based upon the same reasoning. One can also gather the same insight by looking at the boxplots located on the right side of Figure 6.6.3. In short, these results match the graphical interpretations given in Figure 6.6.1.

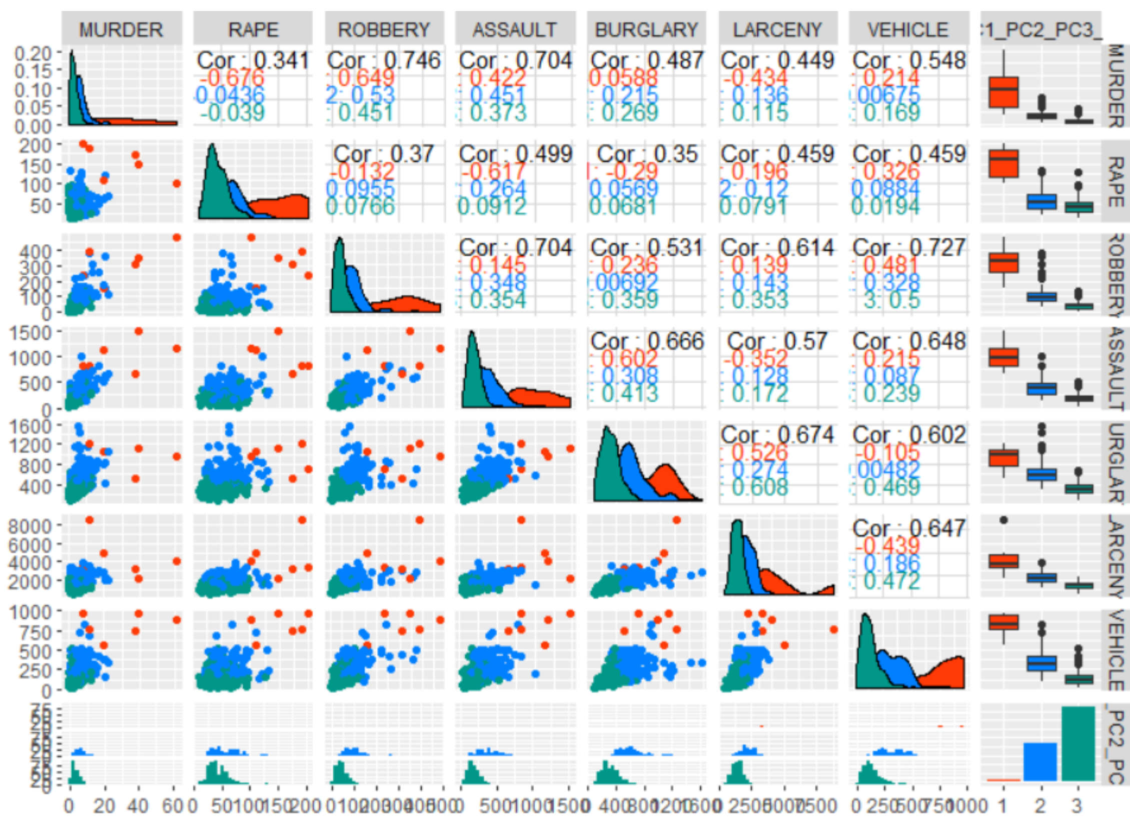


Figure 6.6.4: k-Means, $k = 3$, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, Original Crime 2018

Results from Figure 6.6.4 are analogous to results from Figure 6.6.3.

6.7 Hierarchical Clustering Methods

In a hierarchical clustering algorithm, the data are *not* partitioned into a particular number of clusters at a single step. Instead the clustering consists of a series of partitions, which may run from a single cluster containing all n individuals, to n clusters each containing a single individual. Hierarchical clustering techniques may be subdivided into *agglomerativ*e methods, which proceed by a series of successive fusions of the n individuals into groups, and *divisive* methods, which separate the n individuals successively into smaller groups [12, p. 71].

6.7.1 Agglomerate Clustering Methods

Agglomerative clustering is the most common type of hierarchical clustering used to group objects in clusters based on their similarity. It works in a "bottom-up" manner. That is, each object is initially considered as a single-element cluster (leaf). At each step of the algorithm, the two clusters that are most similar are combined into a new bigger cluster (nodes). This procedure is iterated until all points are members of just one single big cluster (root). The result is a tree-based representation of the fusion of the objects, named a dendrogram [11]. For our analysis, we will focus solely on two agglomerative clustering methods Average and Ward.

6.7.1.1 Average and Ward's Method

Average and Ward's Method can use a Euclidean Distance Matrix $\underset{(n\times n)}{\mathbf{D}}$ as an initial input into the algorithm. Then each method defines a linkage function that takes the distance information $\underset{(n\times n)}{\mathbf{D}}$ and groups pairs of objects into clusters based on some type of similarity criterion. Next, these newly formed clusters are linked to each other to make bigger clusters. This process is iterated until all the objects in the original data set are linked together into a dendrogram.

- Average Linkage Function defines similarity between two clusters as the average distance between the elements in one cluster and the elements in the other cluster.
- Ward's Linkage Function minimizes the total within-cluster variance. At each step the pair of clusters with minimum between-cluster distance are merged.

Note that, at each stage of the clustering process the two clusters, that have the smallest linkage distance, are linked together [11].

6.7.1.1.1 Euclidean Distance Matrix $\underset{(n\times n)}{\mathbf{D_Z}}$ for Standardized Sample $\underset{(n\times p)}{\mathbf{Z}}$

$$\underset{(n\times n)}{\mathbf{D_Z}} = \underset{(n\times n)}{\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{p1} & d_{p2} & \cdots & d_{nn} \end{bmatrix}} = \underset{(n\times n)}{\begin{bmatrix} 0 & d_{12} & \cdots & d_{1n} \\ d_{21} & 0 & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{p1} & d_{p2} & \cdots & 0 \end{bmatrix}}$$

where

$$d_{jl} = d\left(\underset{(p\times 1)}{\mathbf{z}_j}, \underset{(p\times 1)}{\mathbf{z}_l}\right) = \sqrt{\sum_{k=1}^{p}(z_{jk} - z_{lk})^2}$$

$$= \sqrt{(z_{j1} - z_{l1})^2 + (z_{j2} - z_{l2})^2 + \cdots + (z_{jp} - z_{lp})^2}$$

for $j, l = 1,2, \dots, n$.

6.7.1.1.2 Euclidean Distance Matrix $\underset{(n\times n)}{\mathbf{D_{\hat{Y}}}}$ for Sample Principal Components $\underset{(n\times p)}{\widehat{\mathbf{Y}}}$

$$\underset{(n\times n)}{\mathbf{D_{\hat{Y}}}} = \underset{(n\times n)}{\begin{bmatrix} \hat{d}_{11} & \hat{d}_{12} & \cdots & \hat{d}_{1n} \\ \hat{d}_{21} & \hat{d}_{22} & \cdots & \hat{d}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{d}_{p1} & \hat{d}_{p2} & \cdots & \hat{d}_{nn} \end{bmatrix}} = \underset{(n\times n)}{\begin{bmatrix} 0 & \hat{d}_{12} & \cdots & \hat{d}_{1n} \\ \hat{d}_{21} & 0 & \cdots & \hat{d}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{d}_{p1} & \hat{d}_{p2} & \cdots & 0 \end{bmatrix}}$$

where

$$\hat{d}_{jl} = d\left(\underset{(p\times 1)}{\hat{\mathbf{y}}_j}, \underset{(p\times 1)}{\hat{\mathbf{y}}_l}\right) = \sqrt{\sum_{k=1}^{p}(\hat{y}_{jk} - \hat{y}_{lk})^2}$$

$$= \sqrt{(\hat{y}_{j1} - \hat{y}_{l1})^2 + (\hat{y}_{j2} - \hat{y}_{l2})^2 + \cdots + (\hat{y}_{jp} - \hat{y}_{lp})^2}$$

for $j, l = 1,2, \dots, n$.

6.7.1.1.3 Average and Ward's Clustering Pseudo-Code

- Prepare the sample data.
- Compute the Euclidean distance matrix $\underset{(n\times n)}{\mathbf{D}}$.
- Use linkage function to group objects into dendrogram based on $\underset{(n\times n)}{\mathbf{D}}$.
- Determine where to partition the dendrogram branches, creating k clusters

[11].

6.7.2 Euclidean Distance Matrices

6.7.2.1 Euclidean Distance Matrix for Standardized US Crime 2018

Table 6.7.1: Euclidean Distance Matrix for Standardized US Crime 2018, First Five Observations

Euclidean Distance Matrix for Standardized US Crime 2018, First Five Observations					
	Abilene	**Akron**	**Albany_GA**	**Albany_NY**	**Albuquerque**
Abilene	0	0.8	3.1	1.9	6.1
Akron	0.8	0	3.2	1.3	6.3
Albany_GA	3.1	3.2	0	4.2	4.6
Albany_NY	1.9	1.3	4.2	0	7.3
Albuquerque	6.1	6.3	4.6	7.3	0

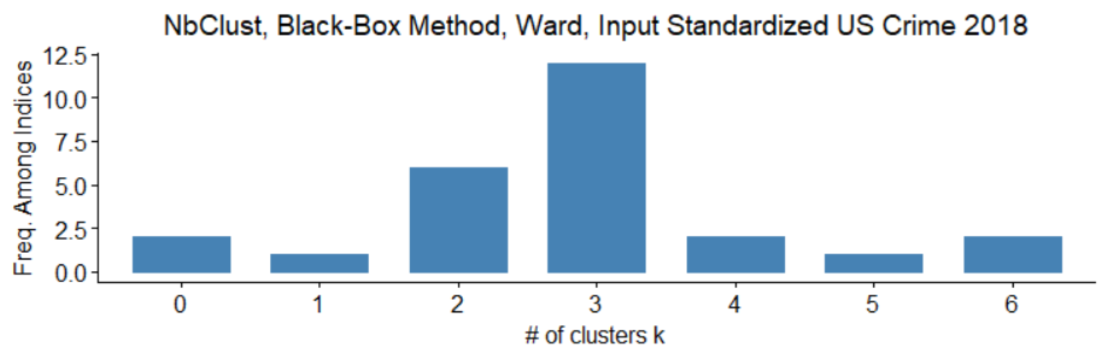
6.7.2.2 Euclidean Distance Matrix for $\hat{y}_1, \hat{y}_2, \hat{y}_3$

Table 6.7.2: Euclidean Distance Matrix for $\hat{y}_1, \hat{y}_2, \hat{y}_3$, First Five Observations

Euclidean Distance Matrix for y1, y2, y3, First Five Observations					
	Abilene	**Akron**	**Albany_GA**	**Albany_NY**	**Albuquerque**
Abilene	0	0.7	2.8	1.5	5.5
Akron	0.7	0	2.9	1.1	5.8
Albany_GA	2.8	2.9	0	3.8	3.4
Albany_NY	1.5	1.1	3.8	0	6.9
Albuquerque	5.5	5.8	3.4	6.9	0

6.7.3 Wards Method

6.7.3.1 Choosing *k*



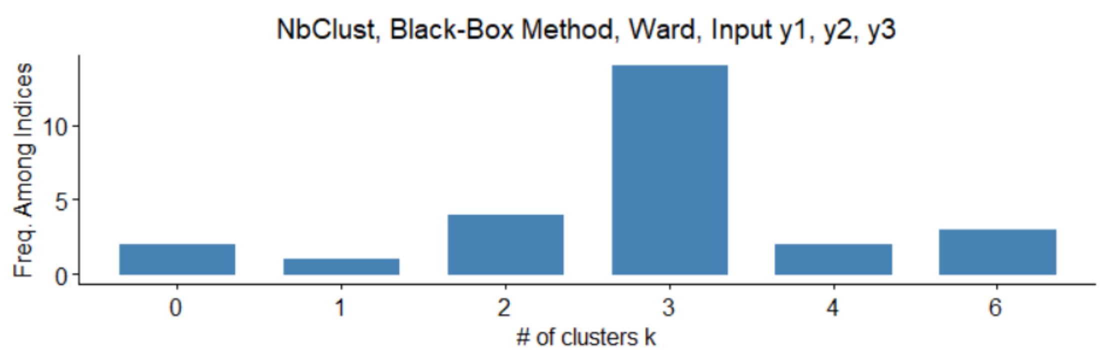


Figure 6.7.1: NbClust, Black-Box Method, Ward

In Figure 6.7.1, the optimal number of clusters is found to be $k = 3$ for both inputs, Standardized US Crime 2018 and $\hat{y}_1, \hat{y}_2, \hat{y}_3$. We will continue our analysis with $k = 3$.

6.7.3.2 Ward, $k = 3$

6.7.3.2.1 Ward, $k = 3$, Cluster Sizes

Table 6.7.3: Ward, $k = 3$, Cluster Sizes

k=3, Ward, Standardized Crime 2018, Cluster Size			
cluster	**1**	**2**	**3**
size	11	48	268
k=3, Ward, y1, y2, y3, Cluster Size			
cluster	**1**	**2**	**3**
size	6	72	249

Interestingly, one can see that the cluster sizes for Wards algorithm, in Table 6.7.3, match the cluster sizes in the k-Means algorithm, for cluster 1 (Table 6.6.1). That is, cluster 1 has 11 members for k-Means and Wards methods, with respect to input standardized Crime 2018. In the same way, cluster 1 has 6 members for k-Means and Wards methods, with respect to input $\hat{y}_1, \hat{y}_2, \hat{y}_3$.Yet, the other assignments for Wards are not the same as for k-Means. At first glance, it looks like Wards method produces larger cluster 3's and smaller cluster 2's then in the k-Means analysis.

6.7.3.2.2 Ward, $k = 3$, Difference in Cluster Assignments

Table 6.7.4: Ward, $k = 3$, Differences in Cluster Assignments

Ward, k=3, Differences in Cluster Assignments				
	y1, y2, y3			
Standardized Crime 2018	**cluster**	**1**	**2**	**3**
	1	6	5	0
	2	0	45	3
	3	0	22	246

Table 6.7.5: Ward, $k = 3$, Differences in Cluster Assignments for Clusters 1,2,3

Ward, k=3, Differences in Assignments for Cluster 1 and 2		
Metropolitan Statistical Area	**Standardized Crime 2018**	**y1, y2, y3**
Albuquerque	1	2
Chicago	1	2
Houston	1	2
Memphis	1	2
Nashville	1	2
Ward, k=3, Differences in Assignments for Cluster 2 and 3		
Metropolitan Statistical Area	**Standardized Crime 2018**	**y1, y2, y3**
Dothan	2	3
Jackson_TN	2	3
Lafayette_LA	2	3
Ward, k=3, Differences in Assignments for Cluster 3 and 2		
Metropolitan Statistical Area	**Standardized Crime 2018**	**y1, y2, y3**
Battle_Creek	3	2
Billings	3	2
Chattanooga	3	2
Cleveland	3	2
Colorado_Springs	3	2
Columbia_SC	3	2
Farmington	3	2
Fresno	3	2
Gainesville_FL	3	2
Gulfport	3	2
Jackson_MI	3	2
Medford	3	2
Modesto	3	2
Muskegon	3	2
Panama_City	3	2
Rapid_City	3	2
Salt_Lake	3	2
San_Francisco	3	2
Seattle	3	2
Stockton	3	2
Tuscaloosa	3	2
Warner_Robins	3	2

6.7.3.2.3 Ward, $k = 3$, Sample Mean Vectors

Table 6.7.6: Ward, $k = 3$, Sample Cluster Mean Vectors

Ward, k=3, Standardized Crime 2018, Sample Cluster Mean Vectors							
cluster	**MURDER**	**RAPE**	**ROBBERY**	**ASSAULT**	**BURGLARY**	**LARCENY**	**VEHICLE**
1	22.7091	111.782	312.0364	865.3273	811.3909	3678.564	681.509
2	8.33125	64.8854	103.6271	439.1646	783.5458	2451.292	348.371
3	3.81493	46.0082	54.68246	215.4007	358.2201	1543.24	172.183
Ward, k=3, y1, y2, y3, Sample Cluster Mean Vectors							
cluster	**MURDER**	**RAPE**	**ROBBERY**	**ASSAULT**	**BURGLARY**	**LARCENY**	**VEHICLE**
1	29.55	153.2	316.9833	1010	925.7	4402.25	819.6
2	7.72639	65.3292	113.8458	434.8806	703.2458	2419.354	359.553
3	3.76908	44.3831	52.05863	204.6365	346.7896	1490.392	158.868
Original Sample Mean Vector for Crime 2018							
	MURDER	**RAPE**	**ROBBERY**	**ASSAULT**	**BURGLARY**	**LARCENY**	**VEHICLE**
	5.11	50.99	70.52	270.11	435.9	1748.36	215.18

Table 6.7.6 shows that the cluster sample means for Wards method are not very different than those of the k-Means (Table 6.6.4). Thus, for both inputs we can again label cluster 3 the *below average crime cluster*; cluster 2 the *above average crime cluster*; and cluster 1 the *extreme crime cluster*. Despite that similarity to the k-Means, there are some key differences. First, the cluster 3 sample means for input Standardized Crime 2018 and input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ are larger than the cluster 3 samples means from the k-Means analysis (Table 6.6.4). Second, the cluster 2 sample means for input Standardized Crime 2018 and input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ are larger than the cluster 2 samples means from the k-Means analysis (Table 6.6.4). This can be visualized later using the cluster assignments plotted on $\hat{y}_1, \hat{y}_2, \hat{y}_3$. Third and finally, the sample mean components for input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, in Wards, are not systematically larger than the sample mean components for input Standardized Crime 2018; as they were with k-Means (Table 6.6.4).

6.7.3.2.4 Ward, $k = 3$, Rectangular Dendrograms

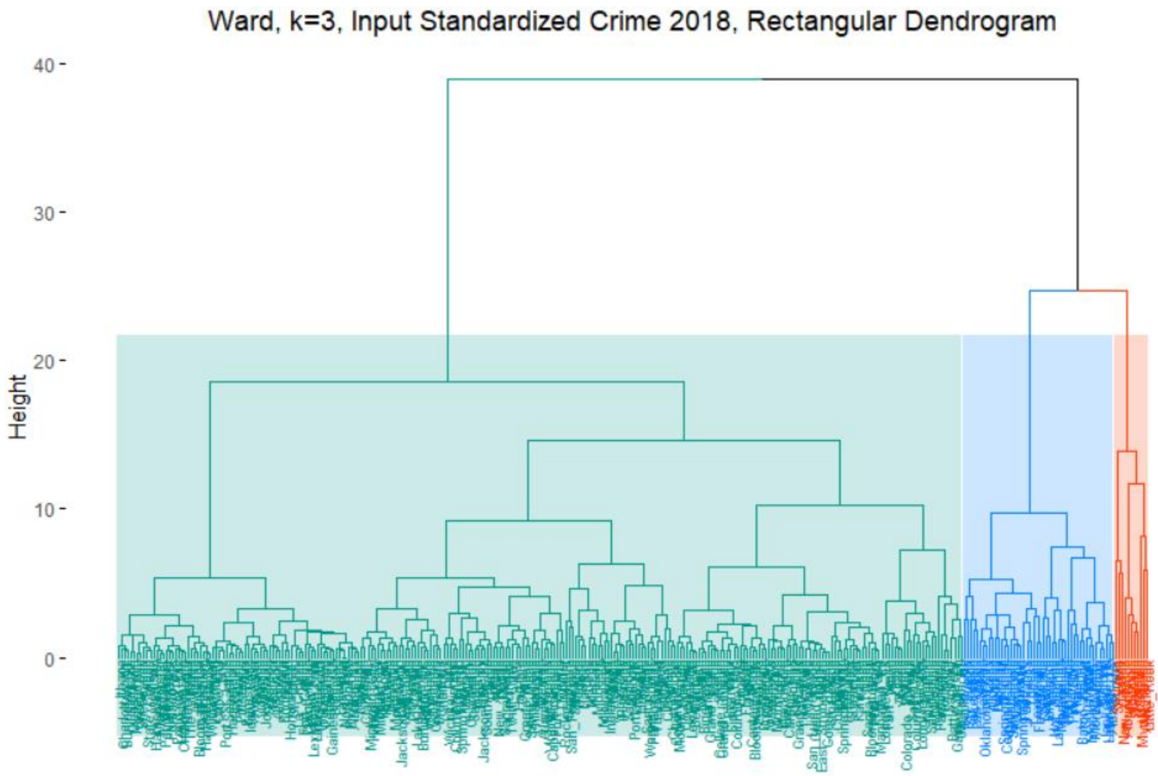


Figure 6.7.2: Ward, $k = 3$, Input Standardized Crime 2018, Rectangular Dendrogram

In the dendrogram displayed above, Figure 6.7.2, each leaf corresponds to a metropolitan statistical area. As we move up the tree, areas that are similar to each other are combined into branches, which are themselves fused at a higher height. The height of the fusion, provided on the vertical axis, indicates the similarity/distance between the two objects/clusters. The higher the height of the fusion, the less similar the objects/clusters are [11].

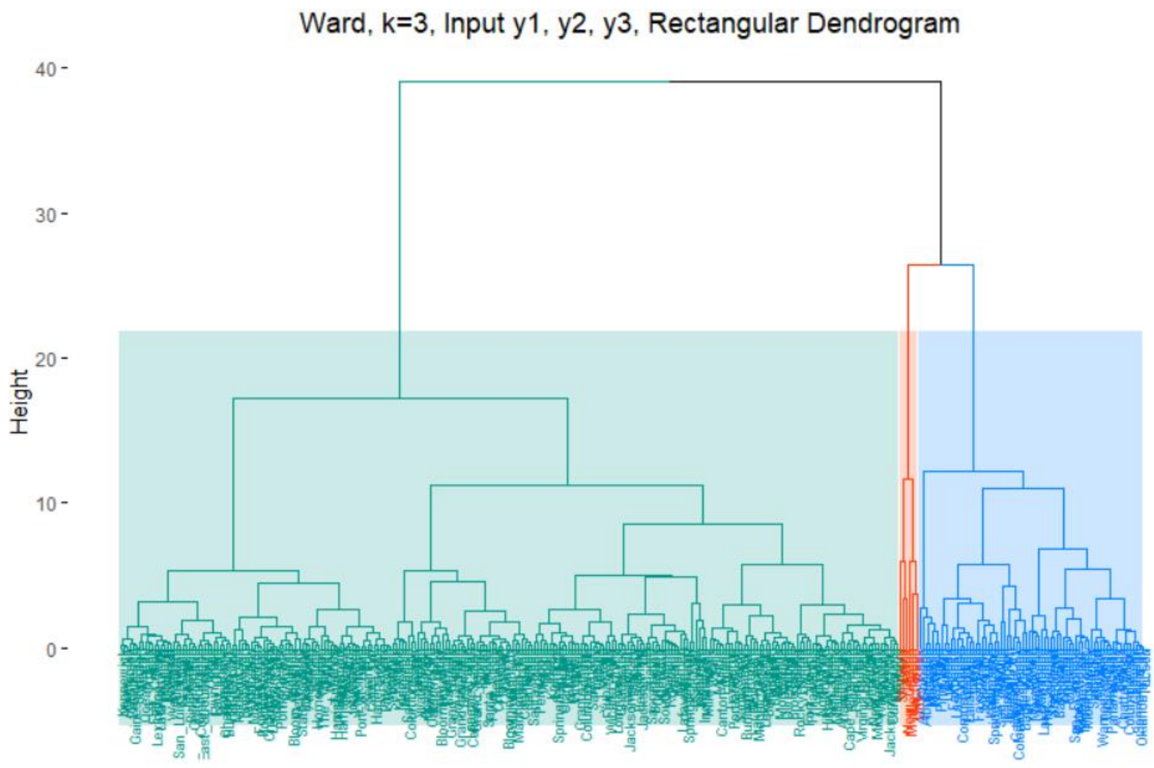


Figure 6.7.3: Ward, $k = 3$, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, Rectangular Dendrogram

Comparing Figure 6.7.3 to Figure 6.7.4, one can visually see that for input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, Wards method produces a larger cluster 2 and a smaller cluster 3.

6.7.3.2.5 Ward, $k = 3$, Scatterplots on $\hat{y}_1, \hat{y}_2, \hat{y}_3$

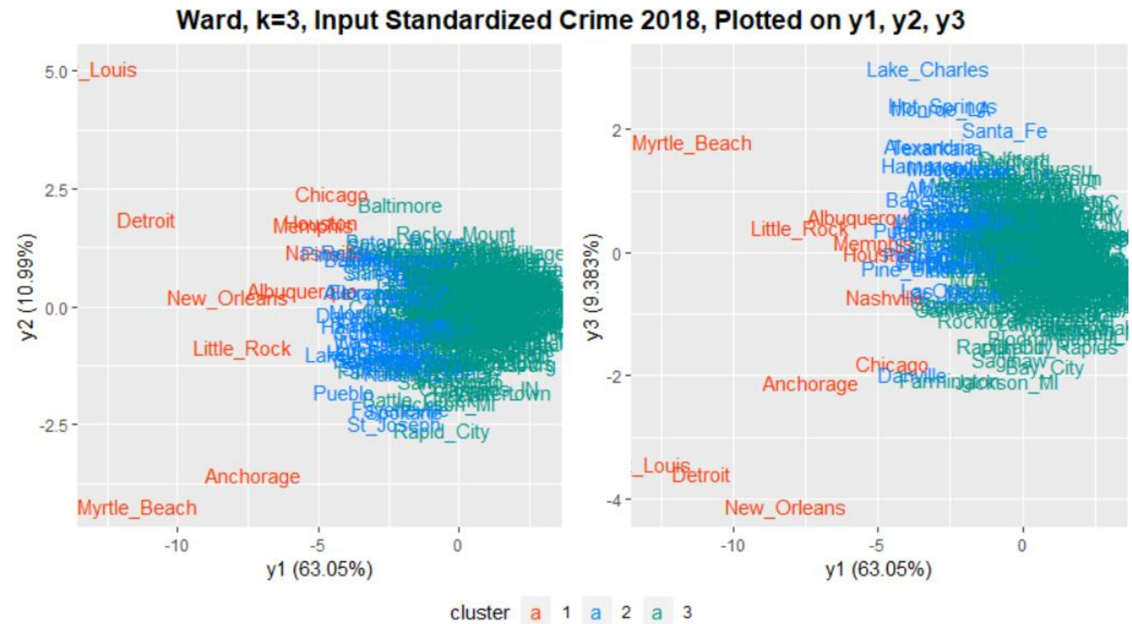


Figure 6.7.4: Ward, $k = 3$, Input Standardized Crime 2018, Plotted on $\hat{y}_1, \hat{y}_2, \hat{y}_3$

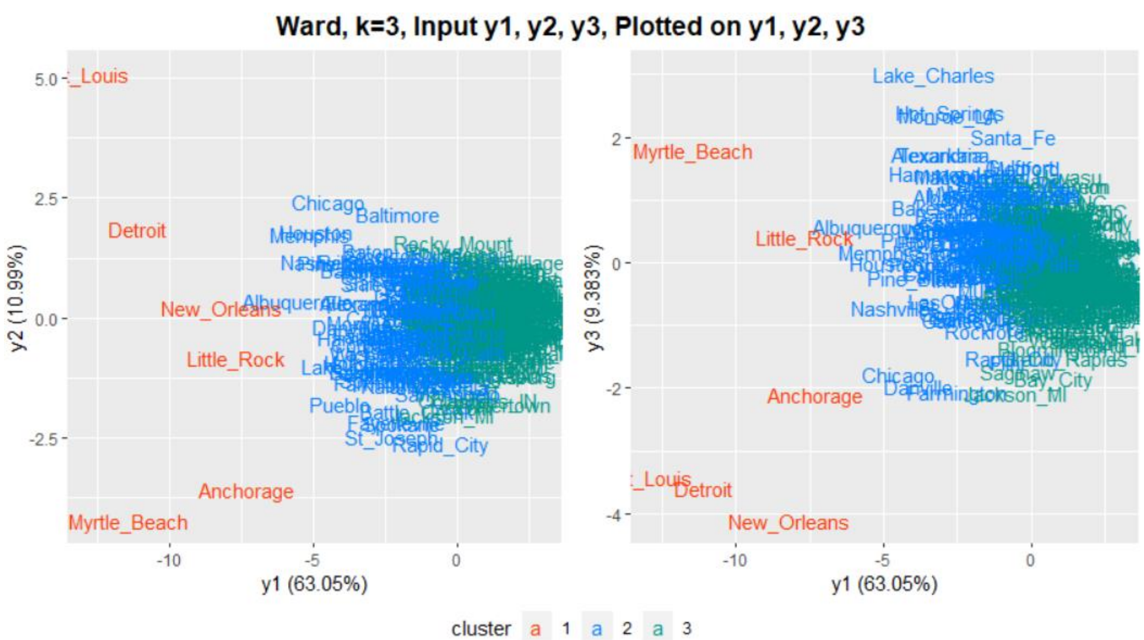


Figure 6.7.5: Ward, $k = 3$, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, Plotted on $\hat{y}_1, \hat{y}_2, \hat{y}_3$

Referring to Figure 6.7.4 and Figure 6.7.5, one can see that cluster 3 for input Standardized Crime 2018 and input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ have become larger compared to their k-Means counterparts in Figure 6.6.1 and Figure 6.6.2. Therefore, using Wards algorithm, cluster 3's centroids, on the $\hat{y}_1$ axis, have shifted to the left. Since $\hat{y}_1$ is the *general crime component*, shifting the cluster 3's to the left, causes the sample mean components in Table 6.7.6 to increase. This is because the $\hat{y}_1$ eigenvector components are negative; consequently, areas with larger crime rates will have more negative scores on $\hat{y}_1$.

Cluster 2, for input Standardized Crime 2018 and input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ have become smaller compared to their k-Means counterparts in Figure 6.6.1 and Figure 6.6.2. Since, cluster 2 lost metropolitan statistical areas further to the right with respect to the $\hat{y}_1$ dimension, the *general crime component*, the sample mean components in Table 6.7.6 have also increasing. That is, cluster two lost areas with lower crime rates to cluster 3. Hence, the sample mean components increase in the original Crime 2018 dimensions.

6.7.3.2.6 Ward, $k = 3$, Scatterplots on Original Crime 2018 Dimensions

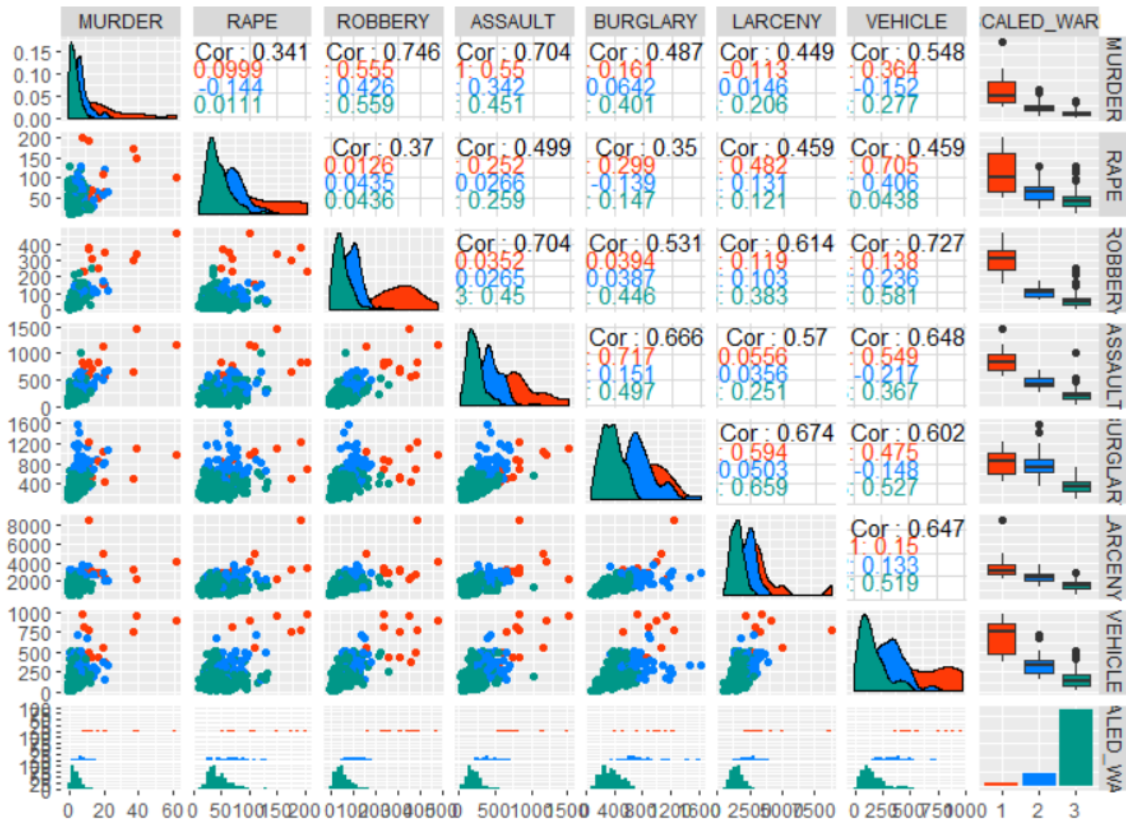


Figure 6.7.6: Ward, $k = 3$, Input Standardized Crime 2018, Original Crime 2018

For Figure 6.7.6, once can verify that cluster 1, has the largest sample means, cluster 2, has the second largest sample means, and cluster 3, has the smallest cluster means.

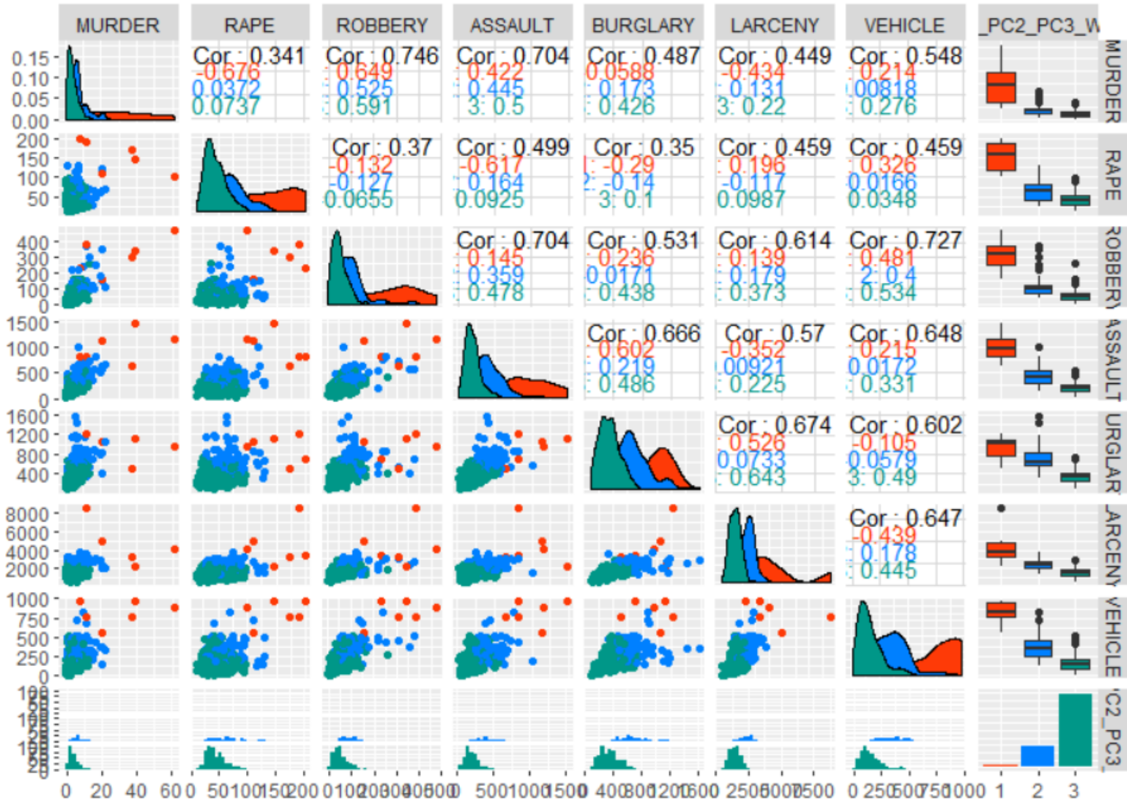


Figure 6.7.7: Ward, $k = 3$, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, Original Crime 2018

For Figure 6.7.7, once can verify that cluster 1, has the largest sample means, cluster 2, has the second largest sample means, and cluster 3, has the smallest cluster means.

6.7.4 Average Method

6.7.4.1 Choosing k

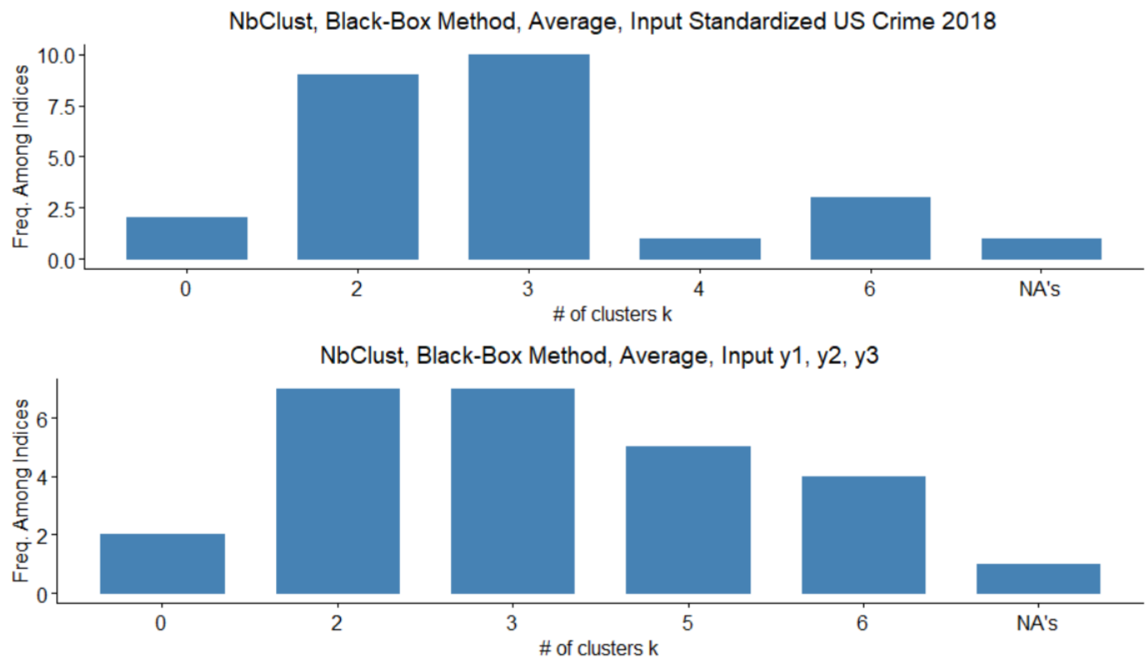


Figure 6.7.8: NbClust, Black-Box Method, Average

In Figure 6.7.8, the optimal number of clusters is found to be $k = 3$ for input Standardized US Crime 2018 and $k = 2, 3$ for input $\hat{y}_1, \hat{y}_2, \hat{y}_3$. We will continue our analysis with $k = 3$.

6.7.4.2 Average, $k = 3$

6.7.4.2.1 Average, $k = 3$, Cluster Sizes

Table 6.7.7: Average, $k = 3$, Cluster Sizes

k=3, Average, Standardized Crime 2018, Cluster Size			
cluster	**1**	**2**	**3**
size	3	3	321
k=3, Average, y1, y2, y3, Cluster Size			
cluster	**1**	**2**	**3**
size	3	3	321

One can see from Table 6.7.7 that the cluster sizes are the same for clusters 1, 2, and 3. We shall see that each cluster also contains the same metropolitan statistical areas. Therefore, there are no differences in cluster assignments for Average, $k = 3$.

6.7.4.2.2 Average, $k = 3$, Sample Mean Vectors

Table 6.7.8: Average, $k = 3$, Sample Cluster Mean Vectors

Average, k=3, Standardized Crime 2018, Sample Cluster Mean Vectors							
cluster	**MURDER**	**RAPE**	**ROBBERY**	**ASSAULT**	**BURGLARY**	**LARCENY**	**VEHICLE**
1	45.63333	139.9	374.9	1096.767	863.5	3190.1	870.9667
2	13.46667	166.5	259.0667	923.2333	987.9	5614.4	768.2333
3	4.656698	49.081	65.91745	256.2801	426.7424	1698.757	203.881
Average, k=3, y1, y2, y3, Sample Cluster Mean Vectors							
cluster	**MURDER**	**RAPE**	**ROBBERY**	**ASSAULT**	**BURGLARY**	**LARCENY**	**VEHICLE**
1	45.63333	139.9	374.9	1096.767	863.5	3190.1	870.9667
2	13.46667	166.5	259.0667	923.2333	987.9	5614.4	768.2333
3	4.656698	49.081	65.91745	256.2801	426.7424	1698.757	203.881
Original Sample Mean Vector for Crime 2018							
	MURDER	**RAPE**	**ROBBERY**	**ASSAULT**	**BURGLARY**	**LARCENY**	**VEHICLE**
	5.11	50.99	70.52	270.11	435.9	1748.36	215.18

Looking at Table 6.7.8 one can see that cluster 3's sample mean vector components, using Average method, have very similar values to the original Crime 2018 sample mean vector components. After all, cluster 3 has 321/327 of the metropolitan statistical areas in its cluster. One could label cluster 3 as the *average crime cluster* even though it is likely composed of places with low, medium, and high crime rates. Cluster 2 and 3 are a bit harder to precisely name. It is clear that, cluster 2 and cluster 3 have larger sample mean components then cluster 1. Although, one can say, cluster 1 has the largest sample mean components on murder, robbery, assault, and vehicle theft. Whereas, cluster 2 has the largest sample mean components on rape, burglary, and larceny. It would be convenient if the clusters were split by crime type, but this is not the case.

6.7.4.2.3 Average, $k = 3$, Rectangular Dendrograms

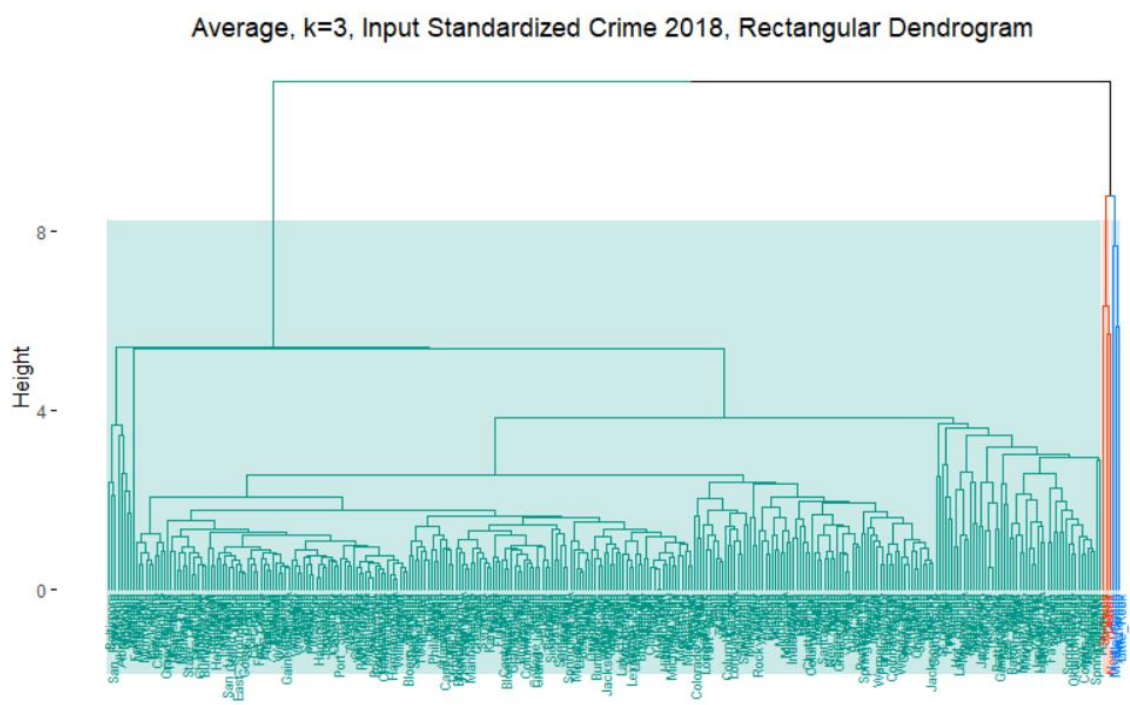


Figure 6.7.9: Average, $k = 3$, Input Standardized Crime 2018, Rectangular Dendrogram

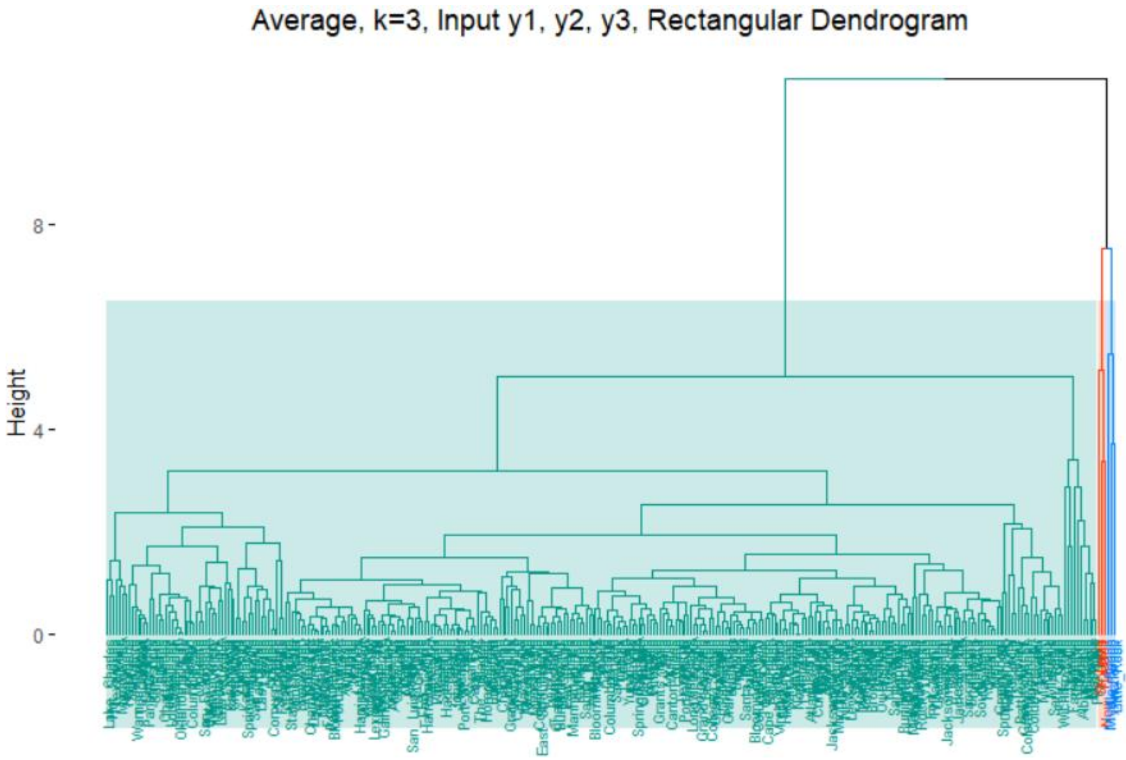


Figure 6.7.10: Average, $k = 3$, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, Rectangular Dendrogram

After reviewing Figure 6.7.9 and Figure 6.7.10, one can see that even though the dendrograms have the same cluster assignments for $k = 3$, they do *not* have identical tree structure. Undoubtedly, if one would increase k (increase the number of clusters), the cluster assignments would change for input Standardized Crime 2018 compared to input $\hat{y}_1, \hat{y}_2, \hat{y}_3$. In Section 6.7.5, we will analytically compare all combinations of dendrograms with respect to inputs and algorithms.

6.7.4.2.4 Average, $k = 3$, Scatterplots on $\hat{y}_1, \hat{y}_2, \hat{y}_3$

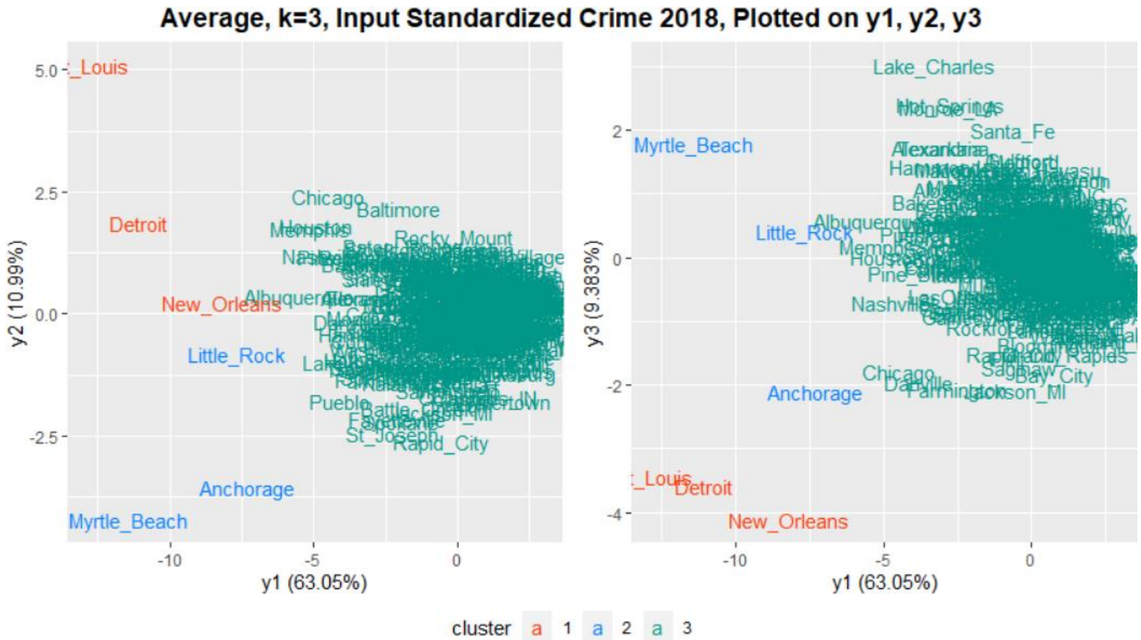


Figure 6.7.11: Average, $k = 3$, Input Standardized Crime 2018, Plotted on $\hat{y}_1, \hat{y}_2, \hat{y}_3$

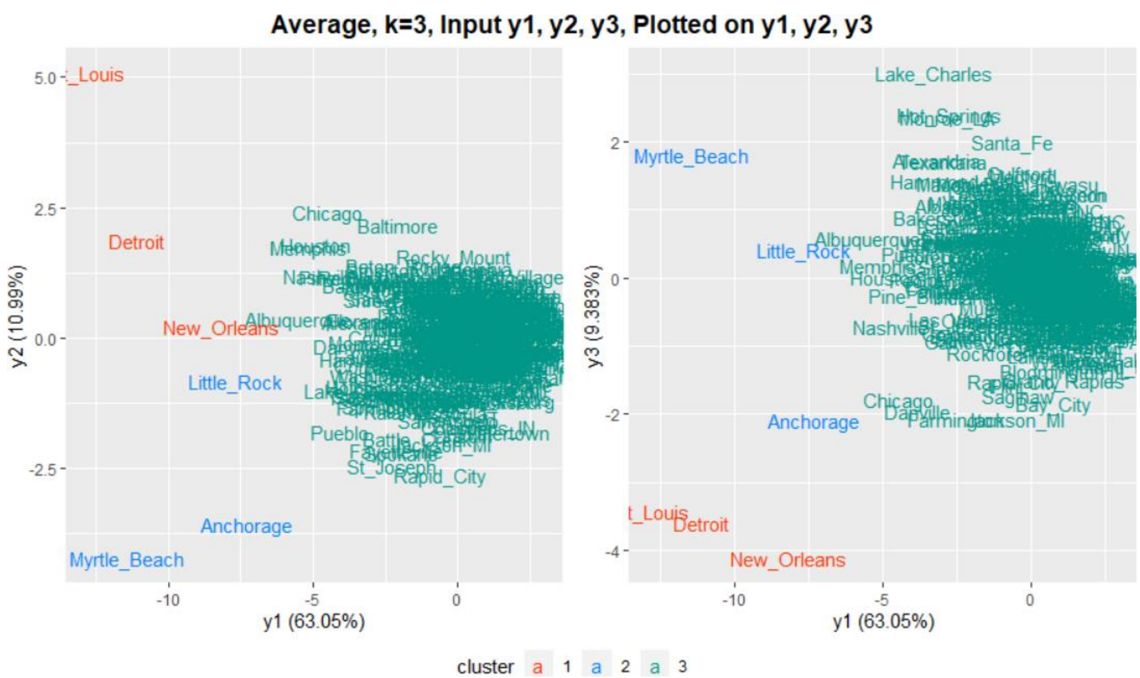


Figure 6.7.12: Average, $k = 3$, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, Plotted on $\hat{y}_1, \hat{y}_2, \hat{y}_3$

As has been noted, cluster assignments for both inputs are same when $k = 3$. Hence, Figure 6.7.11 and Figure 6.7.12 are indistinguishable. Cluster 1 has metropolitan statistical areas St. Louis, Detroit, and New Orleans. Cluster 2 has metropolitan statistical areas Myrtle Beach, Anchorage, and Little Rock.

Looking at the left plot $\hat{y}_2 \sim \hat{y}_1$, one can see cluster 1 sits in the upper left region. In terms of $\hat{y}_1$ (*general crime component*), we know these areas have been classified as having extremely high crime rates. In terms of $\hat{y}_2$ (*heinous crime component*), we know that these areas will have higher murder rates relative to rape rapes because the component is dominated by a negative eigenvector coefficient for rape and a positive eigenvector coefficient for murder. From Figure 6.2.2 (Murder Outliers), one can see that St. Louis, Detroit, and New Orleans have the largest murder rates of the sample in descending order. One should note that in Figure 6.2.4 (Rape Outliers), New Orleans ranks third. Therefore, New Orleans is being pulled back down in the $\hat{y}_2$ direction. Nevertheless, we could cautiously call cluster 1, the *murder cluster*.

Continuing to look at the left plot $\hat{y}_2 \sim \hat{y}_1$, one can see cluster 2 sits in the lower left region. In terms of $\hat{y}_1$, we also know these areas have been classified as having extremely high crime rates. In terms of $\hat{y}_2$, we know that these areas will have higher rape rates relative to murder rapes. This is certainly true for Anchorage and Myrtle Beach because they have the highest rape rates, in descending order, according to Figure 6.2.4. Little Rock, however, has large crime rates on murder and rape; thus, it's getting pulled up in the $\hat{y}_2$ direction. Regardless, one could label cluster 2, the *rape cluster.*

6.7.4.2.5 Average, $k = 3$, Scatterplots on Original Crime 2018 Dimensions

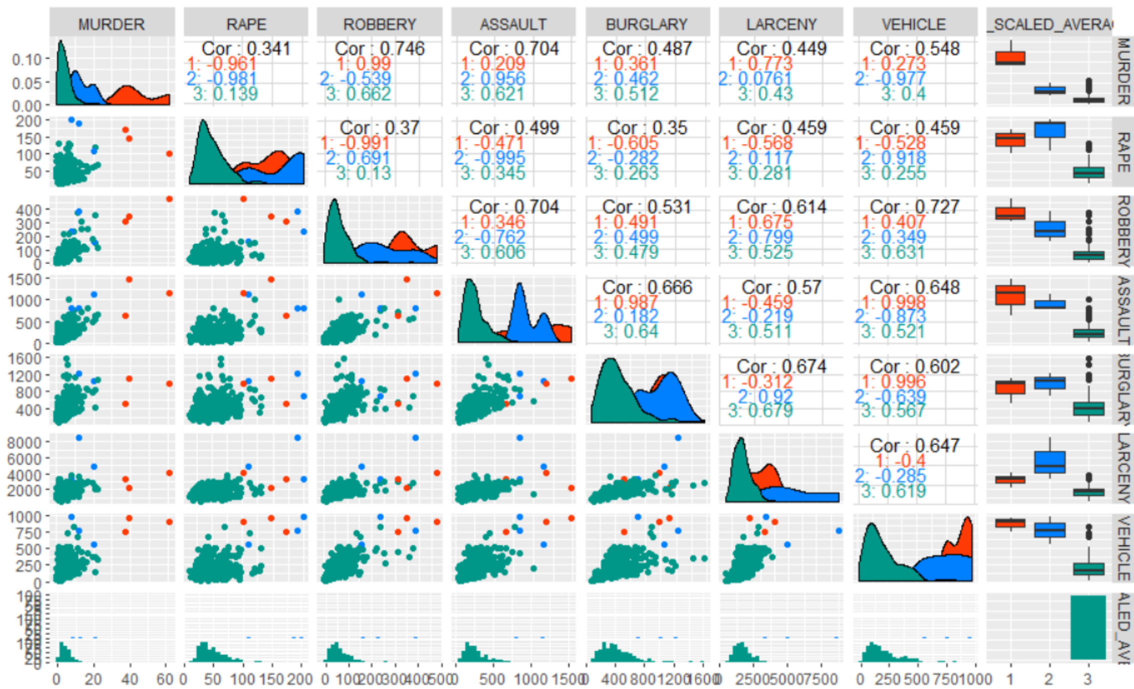


Figure 6.7.13: Average, $k = 3$, Input Standardized Crime 2018, Original Crime 2018

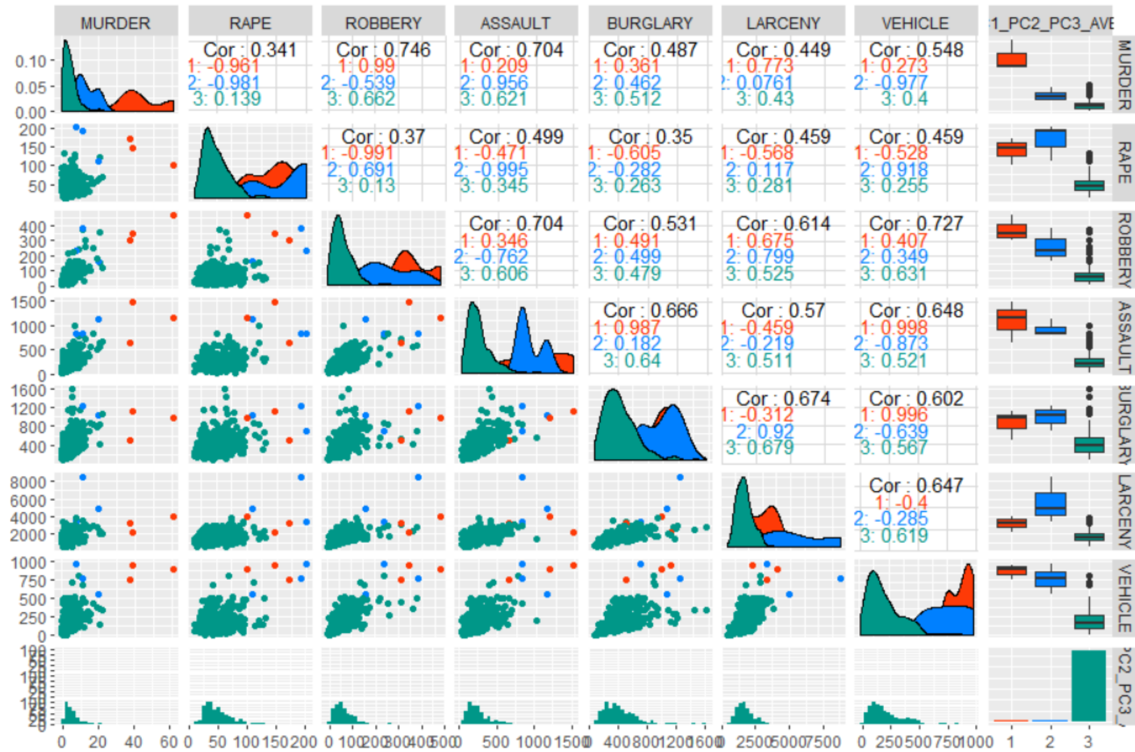


Figure 6.7.14: Average, $k = 3$, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, Original Crime 2018

From Figure 6.7.13 and Figure 6.7.14, once can see cluster 1 has the largest sample mean components on murder, robbery, assault, and vehicle theft. While, cluster 2 has the largest sample mean components on rape, burglary, and larceny (as seen in Table 6.7.8).

6.7.5 Comparing Ward and Average Dendrograms Using Tanglegrams

To visually compare two dendrograms, we'll use the *tanglegram* function (in the R dendextend package), which plots two dendrograms, side by side, with their labels connected by lines. Colored lines represent common subtrees between the two dendrograms, and dashed lines represent unique branches (not common to both trees).

6.7.5.1 Ward, Input S. Crime 2018 vs. Ward, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$

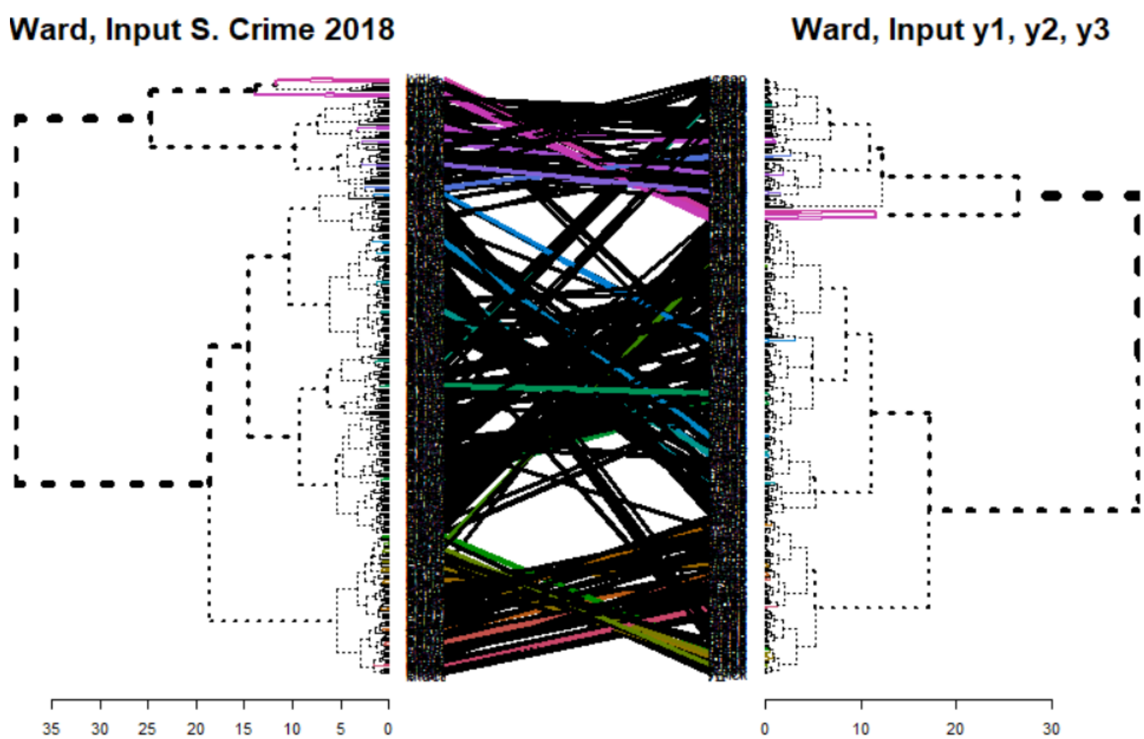


Figure 6.7.15: Ward, Input S. Crime 2018 vs. Ward, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$

The tanglegram for Ward input S. Crime 2018 vs. Ward, input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, in Figure

6.7.15 shows a few common lower initial subtrees where all outer branches are unique. Thus, different input on same algorithm gives very unique dendrograms in this analysis.

6.7.5.2 Average, Input S. Crime 2018 vs. Average, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$

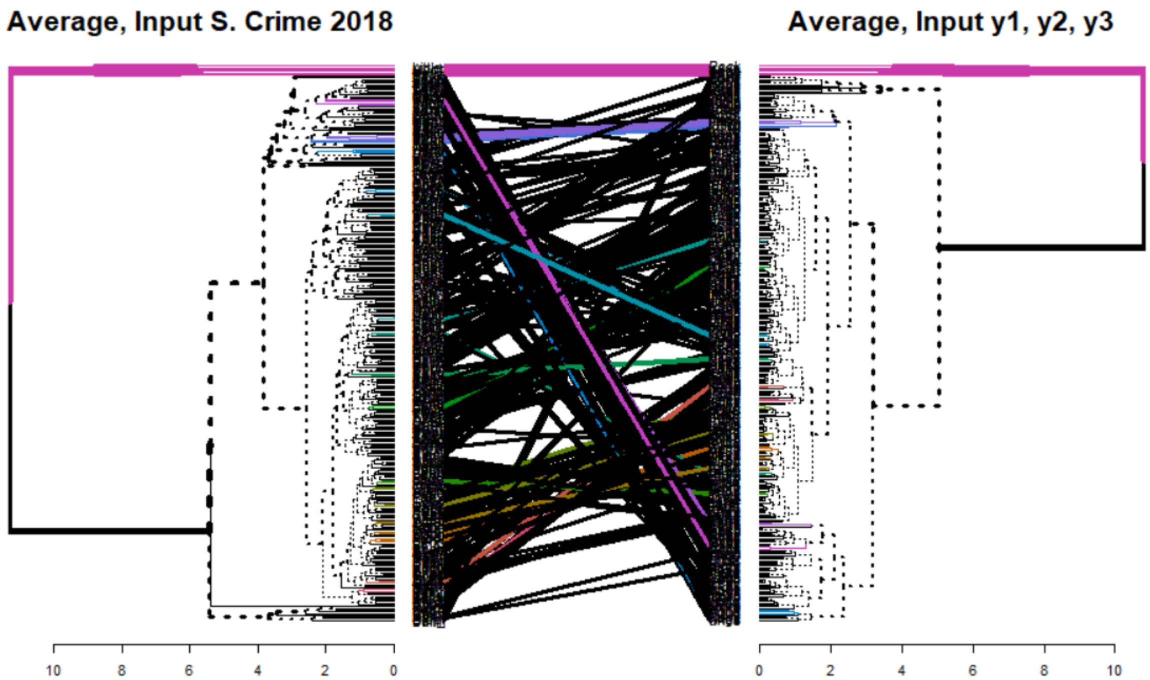


Figure 6.7.16: Average, Input S. Crime 2018 vs. Average, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$

The tanglegram for Average input S. Crime 2018 vs. Average, input $\hat{y}_1, \hat{y}_2, \hat{y}_3$, in Figure 6.7.16 shows a few more common lower initial subtrees where all outer branches are unique. Nonetheless, different input on same algorithm gives very unique dendrograms in this analysis.

6.7.5.3 Ward, Input S. Crime 2018 vs. Average, Input S. Crime 2018

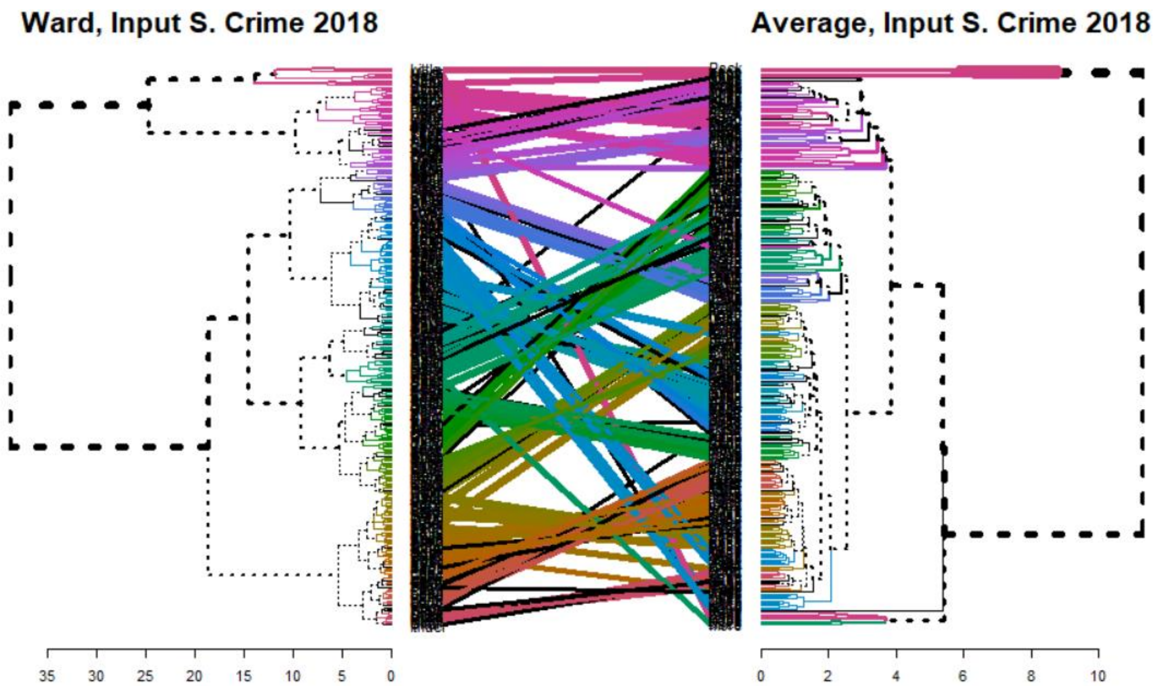


Figure 6.7.17: Ward, Input S. Crime 2018 vs. Average, Input S. Crime 2018

The tanglegram for Ward input S. Crime 2018 vs. Average input S. Crime 2018 in Figure 6.7.17 shows many common lower subtrees where all outer branches are unique. In contrast from the last two tanglegrams, the same inputs on a different algorithm gives very similar dendrograms.

6.7.5.4 Ward, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ vs. Average, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$

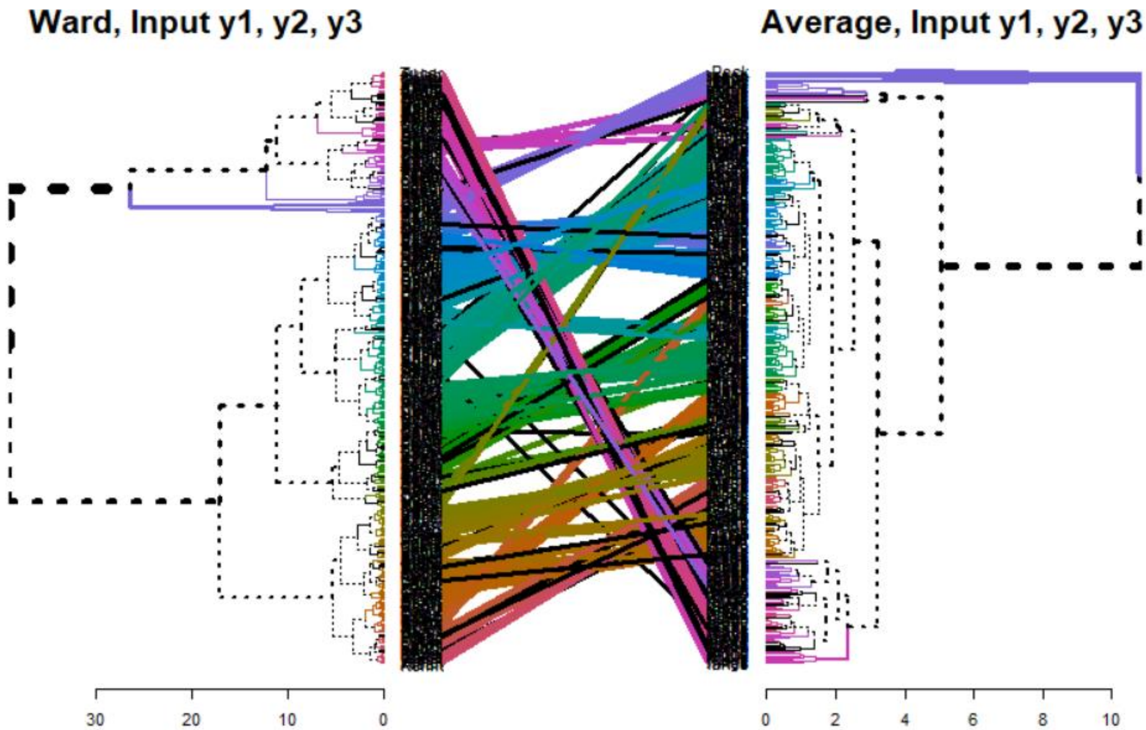


Figure 6.7.18: Ward, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ vs. Average, Input $\hat{y}_1, \hat{y}_2, \hat{y}_3$

The tanglegram for Ward input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ vs. Average, input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ in Figure 6.7.18 shows many common lower subtrees where all outer branches are unique. As one has noted in the previous tanglegram, the same inputs on a different algorithm gives very similar dendrograms.

6.8 Comparison of k-Means, Ward, and Average

We will take a last look at the cluster assignments for k-Means, Ward, and Average methods on $\hat{y}_1, \hat{y}_2$ and compare their respective cluster sizes.

6.8.1 k-Means, Ward, and Average, $k = 3$, Scatterplots on $\hat{y}_1, \hat{y}_2$ and Cluster Sizes

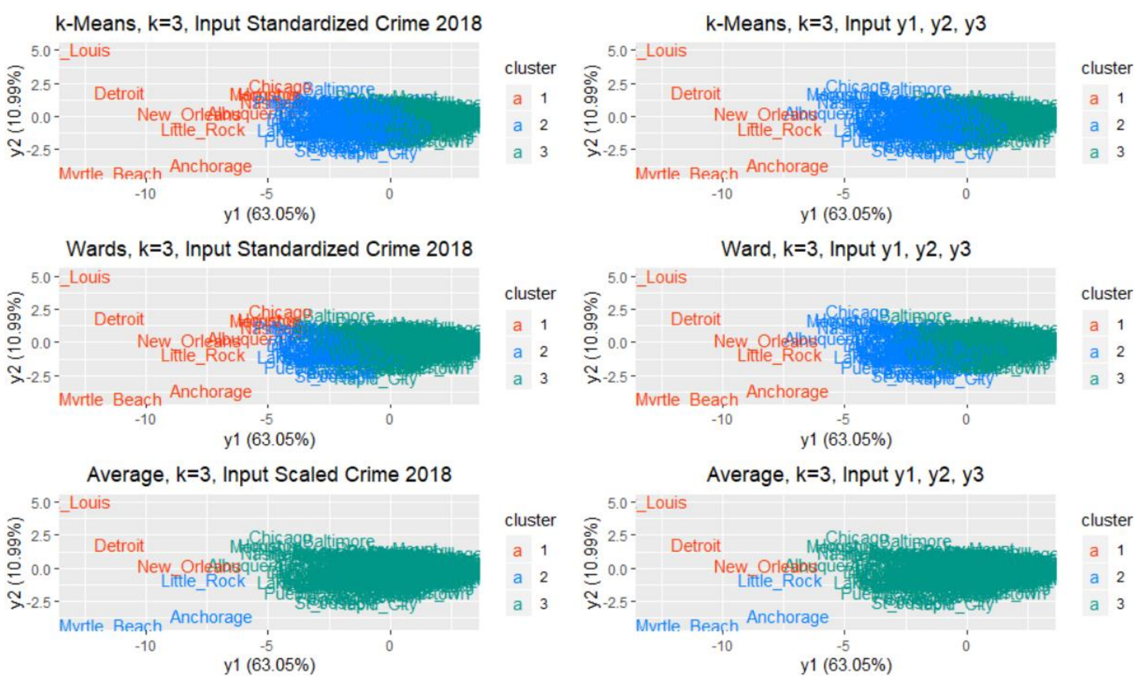


Figure 6.8.1: k-Means, Ward, and Average, $k = 3$, Scatterplots on $\hat{y}_1, \hat{y}_2$

Table 6.8.1: k-Means, Ward, and Average, $k = 3$, Cluster Sizes

k=3, k-Means, Standardized Crime 2018, Cluster Size				k=3, k-Means, y1, y2, y3, Cluster Size			
cluster	**1**	**2**	**3**	**cluster**	**1**	**2**	**3**
size	11	116	200	**size**	6	109	212
k=3, Ward, Standardized Crime 2018, Cluster Size				**k=3, Ward, y1, y2, y3, Cluster Size**			
cluster	**1**	**2**	**3**	**cluster**	**1**	**2**	**3**
size	11	48	268	**size**	6	72	249
k=3, Average, Standardized Crime 2018, Cluster Size				**k=3, Average, y1, y2, y3, Cluster Size**			
cluster	**1**	**2**	**3**	**cluster**	**1**	**2**	**3**
size	3	3	321	**size**	3	3	321

From Figure 6.8.1 and Table 6.8.1, one can see a few general patterns. First, the 6-11 highest crime metropolitan statistical areas are generally in the same cluster, far to the left in the $\hat{y}_1$ direction. With exception of the Average algorithm where the top 6 areas are split by dimension $\hat{y}_2$ (and $\hat{y}_3$ for that matter). That is, cluster 1, is in the upper-left region of $\hat{y}_2 \sim \hat{y}_1$ and cluster 2 is in the lower-left region of $\hat{y}_2 \sim \hat{y}_1$. Next, comparing k-Means and Ward, k-Means cluster sizes for cluster 2 are larger than Ward cluster sizes for cluster 2. Conversely, k-Means cluster sizes for cluster 3 are smaller than Ward cluster sizes for cluster 3. Last, k-Means and Ward are similar insofar as, for input Standardized Crime 2018, they include Albuquerque, Chicago, Houston, Memphis, and Nashville into cluster 1. Further, k-Means and Ward, include Chicago, Houston, Memphis, and Nashville into cluster 2 for input $\hat{y}_1, \hat{y}_2, \hat{y}_3$.

Chapter 7

Conclusion and Future Study

There were several interesting findings when conducting our research on the US Crime 2018 data.

Firstly, many of the extreme univariate outliers also stood out in the scatterplots of the sample principal components $\hat{y}_2 \sim \hat{y}_1$ and $\hat{y}_3 \sim \hat{y}_1$. Next, $\hat{y}_1$, the *general crime component* was a good point estimator for the overall crime in an area because it accounted for 63% of the total variability in the Standardized Crime 2018 data and the eigenvector coefficients had approximately equal magnitude with all negative coefficients. Thus, metropolitan statistical areas with larger crime rates generally were farther to the left in the $\hat{y}_1$ dimension.

Then, we observed k-Means and Ward algorithms clustered areas with extreme crime together, above average crime together, and below average crime together. When viewing these assignments on the sample principal components and the original Crime 2018 dimensions, we also noticed that the 2-d scatters where most dense for the below average crime cluster, less dense for the above average crime cluster, and sparse for the extreme clime cluster. This intuitively makes sense because the univariate crime variables are right skewed, so in 2-d, clusters become less dense as crime increases.

Following this, it was clear when comparing dendrograms for Average and Wards methods, using the same inputs gave remarkably similar tree structures. Meanwhile, when using different inputs on the same algorithm, either Average or Ward, the tree structures were vastly different. This was not expected. Although, one should remember that the input Standardized Crime 2018 was 7 dimensions and the input $\hat{y}_1, \hat{y}_2, \hat{y}_3$ was only 3. As a result, we expect that the general tree structures for agglomerative methods, are more sensitive to dimensionality differences in the distance calculations then in differences in the link function criterions.

Largely, this research uncovered metropolitan statistical areas with extreme crime rates on one or more variables using a combination of univariate and bivariate analysis, principal components, and clustering. However, what this paper did not do, was attempt to try to explain the underlying reasons behind these crime intensities. This is a more nuanced question which necessitates qualitative research along with quantitative research. One would need to conduct interviews with local officials, experts in the area, and people in the community. Also, one would need to research newspaper archives, laws, and get a feel for the culture. Therefore, my future research may be to choose a single metropolitan statistical area and focus on one aspect of crime such as looking at why St. Louis has the highest murder rate in the country or why Myrtle Beach or Anchorage have the highest number of rapes per 100, 000 residents.

Finally, Table 7.1 provides all metropolitan statistical areas that could be of interest for future study that have very high or extremely high crime rates on multiples variables. The 1st, 2nd, and 3rd highest crime rates are highlighted below.

Table 7.1: Metropolitan Statistical Areas of Interest for Future Study

Metropolitan Statistical Areas of Interest for Future Study							
1st Highest Crime Rate, 2nd Highest Crime Rate, 3rd Highest Crime Rate							
METRO	**MURDER**	**RAPE**	**ROBBERY**	**ASSAULT**	**BURGLARY**	**LARCENY**	**VEHICLE**
Albuquerque	9.5	70	238.2	766.9	869.9	2838.9	817.8
Anchorage	8.4	200.1	235.2	819.9	703.4	3342.5	970.9
Baltimore	13.3	38.3	258.4	410.8	399.8	1804.4	266.3
Chicago	20.7	66.1	356.1	563.1	429.8	2379.2	372.6
Detroit	38.9	147.2	344	1477.8	1108.3	2235	961.5
Houston	11.8	53.8	373.6	587	696	2804.6	509.6
Lake_Charles	5.8	63.7	85.8	392.5	1576.1	2852.1	348.4
Little_Rock	20.1	109.4	159.1	1130.5	1043.2	4942.6	562
Memphis	17.2	50.6	254.4	820.3	847.1	2994	430
Myrtle_Beach	11.9	190	382.9	819.3	1217.1	8558.1	771.8
Nashville	13.3	69.9	308.2	721.3	528.3	3034	449
New_Orleans	37.1	171.8	307.5	646.9	511.4	3290.3	755.3
St_Louis	60.9	100.7	473.2	1165.6	970.8	4045	896.1

References

[1] B. Everitt, An R and S-Plus Companion to Multivariate Analysis, London: Springer Science+Business Media, 2005.

[2] U. S. D. o. J. F. B. o. I. Justice, "Crime in the United States, 2018," [Online]. Available: https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/topic-pages/about-cius. [Accessed 18 03 2020].

[3] R. A. Johnson and D. W. Wichern, Applied Multivariate Statistical Analysis, Sixth ed., Upper Saddle River, New Jersey: Pearson, 2019.

[4] O. Bretscher, Linear Algebra with Applications, 5th ed., Upper Saddle River: Pearson, 2013.

[5] D. A. Harville, Matrix Algebra From A Statistician's Perspective, New York: Springer Science+Business Media, LLC, 2008.

[6] M. C. Hogg, Introduction to Mathematical Statistics, 8th ed., Boston: Pearson, 2019.

[7] Wikipedia, "Definiteness_of_a_matrix," 17 3 2020. [Online]. Available: https://en.wikipedia.org/wiki/Definiteness_of_a_matrix. [Accessed 25 3 2020].

[8] Wikipedia, "Log-normal distribution," [Online]. Available: https://en.wikipedia.org/wiki/Log-normal_distribution. [Accessed 3 8 2020].

[9] K. Nosakhere, "Reasons for the high rate of rape in Alaska," Anchorage Press, 2018 5 2018. [Online]. Available: https://www.anchoragepress.com/columnists/reasons-for-the-high-rate-of-rape-in-alaska/article_564f37fa-4e4f-11e8-b893-47cba59bd374.html. [Accessed 9 3 2020].

[10] W. F. C. A. C. Rencher, Methods of Multivariate Analysis, 3rd ed., Hoboken, NJ: John Wiley & Sons, Inc., 2012.

[11] A. Kassamdara, Practical Guide To Cluster Analysis in R: Unsupervised Machine Learning, 2017.

[12] S. L. M. L. D. S. B. S. Everitt, Cluster Analysis, 5th ed., London: John Wiley & Sons, 2011.

Appendix

```
##########################################################
##########################################################

# US CRIME 2018 Metropolitan Statistical Area

# Thesis

###########################################################
###########################################################

# US Crime 2018 features

# univariate distribution analysis

# descriptives

####################

#install.packages("psych", dependencies = TRUE)
library("psych")

describe(CRIME_2018_FEAT)

summary(CRIME_2018_FEAT)

###########################################################
###########################################################
```

```
# US Crime 2018 features

# univariate distribution analysis

# Murder

# density, histogram, boxplot, outliers, lower 2.5%
# (percentile)  crime, qqplots, and shapiro-wilk tests

###################

#install.packages("lattice", dependencies = TRUE)
library("lattice")

#install.packages("gridExtra", dependencies = TRUE)
library("gridExtra")

#install.packages("goft", dependencies = TRUE)
library("goft")

#install.packages("ggplot2", dependencies = TRUE)
library("ggplot2")

#install.packages("magrittr", dependencies = TRUE)
library("magrittr")

#install.packages("ggpubr", dependencies = TRUE)
library("ggpubr")

###################
```

```
# Murder

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

# Density

MURDER_DENSITY <- densityplot(~MURDER, data = US_CRIME_2018,
                              main="Murder Density Plot",
                              col = "#00c9f7")

# Histogram

MURDER_HISTOGRAM <- histogram(x=~MURDER,data=US_CRIME_2018,
                              type="density",
                              main="Murder Histogram",
                              col = "#00c9f7",
                              nint = 50)

# test distribution is normal

shapiro.test(x=US_CRIME_2018$MURDER)
```

```
# Normal QQ-Plot

MURDER_QQ_QNORM <- qqmath(x=~MURDER, data = US_CRIME_2018,
                          distribution = qnorm,
                          prepanel = prepanel.qqmathline,
                          panel = function(x, ...) {
                            panel.qqmathline(x, ...)
                            panel.qqmath(x, ...)
                          },
                          main = "Murder Normal QQ-Plot \n SW-Test p-value = 0",
                          col = "#00c9f7")

# test distribution is log normal

# 0 in distribution (cannot compute statistic)

#install.packages("goft", dependencies = TRUE)
#library("goft")

lnorm_test(x=US_CRIME_2018$MURDER)

# Log-Normal QQ-Plot

MURDER_QQ_QLNORM <- qqmath(x=~MURDER, data = US_CRIME_2018,
                           distribution = qlnorm,
                           prepanel = prepanel.qqmathline,
                           panel = function(x, ...) {
                             panel.qqmathline(x, ...)
                             panel.qqmath(x, ...)
                           },
                           main = "Murder Lognormal QQ-Plot \n SW-Test p-value = NA",
                           col = "#00c9f7")
```

```
################################################

# MURDER

# find outliers based on boxplot

OutVals_murder <- boxplot(US_CRIME_2018$MURDER)$out

which(US_CRIME_2018$MURDER %in% OutVals_murder)

US_CRIME_2018[c(20,  23,  58,  76,  81,  99, 169, 185, 206, 209, 224, 227, 273, 286),]

sub_murder_outlier <- as.data.frame(US_CRIME_2018[c(20,  23,  58,  76,  81,  99, 169,
                                                  185, 206, 209, 224, 227, 273, 286),])

sub_murder_outlier_order <- order(sub_murder_outlier$MURDER,decreasing = TRUE)

sub_murder_outlier[sub_murder_outlier_order,]

OUTLIER_MURDER <- as.data.frame(sub_murder_outlier[sub_murder_outlier_order,])
```

```r
# barplot of outliers for Murder

#install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

OUTLIER_MURDER_BAR <- ggplot(OUTLIER_MURDER, aes(x = reorder(CITY,-MURDER),
                                                  y = MURDER)) +
  geom_bar(fill = "#00c9f7", stat = "identity") +
  geom_text(aes(label = MURDER), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Murder Outliers of Sample") +
  xlab("METRO") + ylab("MURDER per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

###################
```

```r
###################

# Murder 2.5th percentile

quantile(US_CRIME_2018$MURDER, .025)

####################

# ordered at or below 2.5th percentile metro statistical
# areas in terms of murder per 100,000

US_CRIME_2018[US_CRIME_2018$MURDER <= .615,]

sub_by_murder_2_5 <- as.data.frame(US_CRIME_2018[US_CRIME_2018$MURDER <= .615,])

sub_by_murder_2_5_order <- order(sub_by_murder_2_5$MURDER,decreasing = FALSE)

sub_by_murder_2_5[sub_by_murder_2_5_order,]

BOTTOM_MURDER <- as.data.frame(sub_by_murder_2_5[sub_by_murder_2_5_order,])
```

```
#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

# barplot Cities with Murder At or Below 2.5th Percentile in ascending order

BOTTOM_MURDER_BAR <- ggplot(BOTTOM_MURDER , aes(x = reorder(CITY,MURDER), y = MURDER))
  geom_bar(fill = "#00c9f7", stat = "identity") +
  geom_text(aes(label = MURDER), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Murder Lower 2.5% of Sample") +
  xlab("METRO") + ylab("MURDER per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

####################

# Murder barplot with lower 2.5% and outliers

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

MURDER_BAR <- ggarrange(BOTTOM_MURDER_BAR,
                        OUTLIER_MURDER_BAR,
                        ncol = 2,
                        nrow = 1)
MURDER_BAR
```

```
############################################################
############################################################

# US Crime 2018 features

# univariate distribution analysis

# Rape

# density, histogram, boxplot, outliers, lower 2.5%
# (percentile) crime, qqplots, and shapiro-wilk tests

####################

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

#install.packages("gridExtra", dependencies = TRUE)
#library("gridExtra")

#install.packages("goft", dependencies = TRUE)
#library("goft")

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")

#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

####################
```

```
# Rape

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

# Density

RAPE_DENSITY <- densityplot(~RAPE, data = US_CRIME_2018,
                            main="Rape Density Plot",
                            col = "#b88cd1")

# Histogram

RAPE_HISTOGRAM <- histogram(x=~RAPE,data=US_CRIME_2018,
                            type="density",
                            main="Rape Histogram",
                            col = "#b88cd1",
                            nint = 50)

# test distribution is normal

shapiro.test(x=US_CRIME_2018$RAPE)
```

```
# Normal QQ-Plot

RAPE_QQ_QNORM <- qqmath(x=~RAPE, data = US_CRIME_2018,
                        distribution = qnorm,
                        prepanel = prepanel.qqmathline,
                        panel = function(x, ...) {
                          panel.qqmathline(x, ...)
                          panel.qqmath(x, ...)
                        },
                        main = "Rape Normal QQ-Plot \n SW-Test p-value = 0",
                        col = "#b88cd1")

# test distribution is log normal

#install.packages("goft", dependencies = TRUE)
#library("goft")

lnorm_test(x=US_CRIME_2018$RAPE)

# Log-Normal QQ-Plot

RAPE_QQ_QLNORM <- qqmath(x=~RAPE, data = US_CRIME_2018,
                         distribution = qlnorm,
                         prepanel = prepanel.qqmathline,
                         panel = function(x, ...) {
                           panel.qqmathline(x, ...)
                           panel.qqmath(x, ...)
                         },
                         main = "Rape Lognormal QQ-Plot \n SW-Test p-value = 0.164",
                         col = "#b88cd1")
```

```r
# combine 4 plots rape

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(RAPE_DENSITY,
             RAPE_HISTOGRAM,
             RAPE_QQ_QNORM,
             RAPE_QQ_QLNORM,
             ncol=2)

###############################################

# RAPE

# find outliers based on boxplot

OutVals_rape <- boxplot(US_CRIME_2018$RAPE)$out

which(US_CRIME_2018$RAPE %in% OutVals_rape)

US_CRIME_2018[c(10,  24,  76,  81,  95, 134, 169, 203, 209, 238, 241, 285),]

sub_rape_outlier <- as.data.frame(US_CRIME_2018[c(10,  24,  76,  81,  95,
                                                  134, 169, 203, 209,
                                                  238, 241, 285),])

sub_rape_outlier_order <- order(sub_rape_outlier$RAPE,decreasing = TRUE)

sub_rape_outlier[sub_rape_outlier_order,]

OUTLIER_RAPE <- as.data.frame(sub_rape_outlier[sub_rape_outlier_order,])
```

```r
# barplot of outliers for Rape

#install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

OUTLIER_RAPE_BAR <- ggplot(OUTLIER_RAPE, aes(x = reorder(CITY,-RAPE), y = RAPE)) +
  geom_bar(fill = "#b88cd1", stat = "identity") +
  geom_text(aes(label = RAPE), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Rape Outliers of Sample") +
  xlab("METRO") + ylab("RAPE per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

###############

# Rape 2.5th percentile

quantile(US_CRIME_2018$RAPE, .025)

################
```

```r
# at or below 2.5th percentile metro statistical areas in terms of rape per 100,000

US_CRIME_2018[US_CRIME_2018$RAPE <= 19.375,]

sub_by_rape_2_5 <- as.data.frame(US_CRIME_2018[US_CRIME_2018$RAPE <= 19.375,])

sub_by_rape_2_5_order <- order(sub_by_rape_2_5$RAPE,decreasing = FALSE)

sub_by_rape_2_5[sub_by_rape_2_5_order,]

BOTTOM_RAPE <- as.data.frame(sub_by_rape_2_5[sub_by_rape_2_5_order,])

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

# barplot Cities with RAPE At or Below 2.5th Percentile in ascending order

BOTTOM_RAPE_BAR <- ggplot(BOTTOM_RAPE , aes(x = reorder(CITY,RAPE), y = RAPE)) +
  geom_bar(fill = "#b88cd1", stat = "identity") +
  geom_text(aes(label = RAPE), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Rape Lower 2.5% of Sample") +
  xlab("METRO") + ylab("RAPE per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

####################
```

```r
# Rape barplot with lower 2.5% and outliers

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

RAPE_BAR <- ggarrange(BOTTOM_RAPE_BAR,
                      OUTLIER_RAPE_BAR,
                      ncol = 2,
                      nrow = 1)
RAPE_BAR

##############################################################
##############################################################
```

```
# US Crime 2018 features

# univariate distribution analysis

# Robbery

# density, histogram, boxplot, outliers, lower 2.5%
# (percentile) crime, qqplots, and shapiro-wilk tests

####################

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

#install.packages("gridExtra", dependencies = TRUE)
#library("gridExtra")

#install.packages("goft", dependencies = TRUE)
#library("goft")

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")

#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

####################
```

```
# ROBBERY

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

# Density

ROBBERY_DENSITY <- densityplot(~ROBBERY, data = US_CRIME_2018,
                               main="Robbery Density Plot",
                               col = "#7449f7")

# Histogram

ROBBERY_HISTOGRAM <- histogram(x=~ROBBERY,data=US_CRIME_2018,
                               type="density",
                               main="Robbery Histogram",
                               col = "#7449f7",
                               nint = 50)

# test distribution is normal

shapiro.test(x=US_CRIME_2018$ROBBERY)
```

```
# Normal QQ-Plot

ROBBERY_QQ_QNORM <- qqmath(x=~ROBBERY, data = US_CRIME_2018,
                           distribution = qnorm,
                           prepanel = prepanel.qqmathline,
                           panel = function(x, ...) {
                             panel.qqmathline(x, ...)
                             panel.qqmath(x, ...)
                           },
                           main = "Robbery Normal QQ-Plot \n SW-Test p-value = 0",
                           col = "#7449f7")

# test distribution is log normal

#install.packages("goft", dependencies = TRUE)
#library("goft")

lnorm_test(x=US_CRIME_2018$ROBBERY)

# Log-Normal QQ-Plot

ROBBERY_QQ_QLNORM <- qqmath(x=~ROBBERY, data = US_CRIME_2018,
                            distribution = qlnorm,
                            prepanel = prepanel.qqmathline,
                            panel = function(x, ...) {
                              panel.qqmathline(x, ...)
                              panel.qqmath(x, ...)
                            },
                            main = "Robbery Lognormal QQ-Plot \n SW-Test p-value = 0",
                            col = "#7449f7")
```

```
# combine 4 plots robbery

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(ROBBERY_DENSITY,
             ROBBERY_HISTOGRAM,
             ROBBERY_QQ_QNORM,
             ROBBERY_QQ_QLNORM,
             ncol=2)

########################################

# ROBBERY

# find outliers based on boxplot

OutVals_robbery <- boxplot(US_CRIME_2018$ROBBERY)$out

which(US_CRIME_2018$ROBBERY %in% OutVals_robbery)

US_CRIME_2018[c(5, 10, 20, 58, 81, 130, 185, 203, 206, 209, 224, 259, 286, 287),]

sub_robbery_outlier <- as.data.frame(US_CRIME_2018[c(5, 10, 20, 58, 81,
                                                     130, 185, 203, 206,
                                                     209, 224, 259, 286, 287),])

sub_robbery_outlier_order <- order(sub_robbery_outlier$ROBBERY,decreasing = TRUE)

sub_robbery_outlier[sub_robbery_outlier_order,]
```

```r
OUTLIER_ROBBERY <- as.data.frame(sub_robbery_outlier[sub_robbery_outlier_order,])

# barplot of outliers for Robbery

#install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

OUTLIER_ROBBERY_BAR <- ggplot(OUTLIER_ROBBERY, aes(x = reorder(CITY,-ROBBERY),
                                                   y = ROBBERY)) +
  geom_bar(fill = "#7449f7", stat = "identity") +
  geom_text(aes(label = ROBBERY), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Robbery Outliers of Sample") +
  xlab("METRO") + ylab("ROBBERY per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

###############

# Robbery 2.5th percentile

quantile(US_CRIME_2018$ROBBERY, .025)

################
```

```r
# at or below 2.5th percentile metro statistical areas in terms of robbery per 100,000

US_CRIME_2018[US_CRIME_2018$ROBBERY <= 9.73,]

sub_by_robbery_2_5 <- as.data.frame(US_CRIME_2018[US_CRIME_2018$ROBBERY <= 9.73,])

sub_by_robbery_2_5_order <- order(sub_by_robbery_2_5$ROBBERY,decreasing = FALSE)

sub_by_robbery_2_5[sub_by_robbery_2_5_order,]

BOTTOM_ROBBERY <- as.data.frame(sub_by_robbery_2_5[sub_by_robbery_2_5_order,])

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

# barplot Cities with ROBBERY At or Below 2.5th Percentile in ascending order

BOTTOM_ROBBERY_BAR <- ggplot(BOTTOM_ROBBERY , aes(x = reorder(CITY,ROBBERY),
                                                  y = ROBBERY)) +
  geom_bar(fill = "#7449f7", stat = "identity") +
  geom_text(aes(label = ROBBERY), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Robbery Lower 2.5% of Sample") +
  xlab("METRO") + ylab("ROBBERY per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

####################
```

```
# Robbery barplot with lower 2.5% and outliers

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

ROBBERY_BAR <- ggarrange(BOTTOM_ROBBERY_BAR,
                         OUTLIER_ROBBERY_BAR,
                         ncol = 2,
                         nrow = 1)
ROBBERY_BAR
```

```
############################################################
############################################################

# US Crime 2018 features

# univariate distribution analysis

# Assault

# density, histogram, boxplot, outliers, lower 2.5%
# (percentile) crime, qqplots, and shapiro-wilk tests

####################

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

#install.packages("gridExtra", dependencies = TRUE)
#library("gridExtra")

#install.packages("goft", dependencies = TRUE)
#library("goft")

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")

#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

####################
```

```
####################

# Assault

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

# Density

ASSAULT_DENSITY <- densityplot(~ASSAULT, data = US_CRIME_2018,
                               main="Assault Density Plot",
                               col = "#ca3886")

# Histogram

ASSAULT_HISTOGRAM <- histogram(x=~ASSAULT,data=US_CRIME_2018,
                               type="density",
                               main="Assault Histogram",
                               col = "#ca3886",
                               nint = 50)

# test distribution is normal

shapiro.test(x=US_CRIME_2018$ASSAULT)
```

```
# Normal QQ-Plot

ASSAULT_QQ_QNORM <- qqmath(x=~ASSAULT, data = US_CRIME_2018,
                           distribution = qnorm,
                           prepanel = prepanel.qqmathline,
                           panel = function(x, ...) {
                             panel.qqmathline(x, ...)
                             panel.qqmath(x, ...)
                           },
                           main = "Assault Normal QQ-Plot \n SW-Test p-value = 0",
                           col = "#ca3886")

# test distribution is log normal

#install.packages("goft", dependencies = TRUE)
#library("goft")

lnorm_test(x=US_CRIME_2018$ASSAULT)

# Log-Normal QQ-Plot

ASSAULT_QQ_QLNORM <- qqmath(x=~ASSAULT, data = US_CRIME_2018,
                            distribution = qlnorm,
                            prepanel = prepanel.qqmathline,
                            panel = function(x, ...) {
                              panel.qqmathline(x, ...)
                              panel.qqmath(x, ...)
                            },
                            main = "Assault Log-Normal QQ-Plot \n SW-Test p-value = 0.3352",
                            col = "#ca3886")
```

```
# combine 4 plots robbery

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(ASSAULT_DENSITY,
             ASSAULT_HISTOGRAM,
             ASSAULT_QQ_QNORM,
             ASSAULT_QQ_QLNORM,
             ncol=2)

###############################################

# ASSAULT

# find outliers based on boxplot

OutVals_assault <- boxplot(US_CRIME_2018$ASSAULT)$out

which(US_CRIME_2018$ASSAULT %in% OutVals_assault)

US_CRIME_2018[c(3, 5,  6,  10, 12,  76, 81,  95, 120, 130, 169, 174, 185, 203,
                206, 209, 214, 227, 286, 318),]

sub_assault_outlier <- as.data.frame(US_CRIME_2018[c(3, 5,  6,  10, 12,  76, 81,
                                                     95, 120, 130, 169, 174, 185,
                                                     203, 206, 209, 214, 227,
                                                     286, 318),])

sub_assault_outlier_order <- order(sub_assault_outlier$ASSAULT,decreasing = TRUE)

sub_assault_outlier[sub_assault_outlier_order,]

OUTLIER_ASSAULT <- as.data.frame(sub_assault_outlier[sub_assault_outlier_order,])

# barplot of outliers for Assault

#install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

OUTLIER_ASSAULT_BAR <- ggplot(OUTLIER_ASSAULT, aes(x = reorder(CITY,-ASSAULT),
                                                   y = ASSAULT)) +
  geom_bar(fill = "#ca3886", stat = "identity") +
  geom_text(aes(label = ASSAULT), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Assault Outliers of Sample") +
  xlab("METRO") + ylab("ASSAULT per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))


####################

# Assault 2.5th percentile

quantile(US_CRIME_2018$ASSAULT, .025)

####################
```

```
# at or below 2.5th percentile metro statistical areas in terms of assault per 100,000

US_CRIME_2018[US_CRIME_2018$ASSAULT <= 61.59,]

sub_by_assault_2_5 <- as.data.frame(US_CRIME_2018[US_CRIME_2018$ASSAULT <= 61.59,])

sub_by_assault_2_5_order <- order(sub_by_assault_2_5$ASSAULT,decreasing = FALSE)

sub_by_assault_2_5[sub_by_assault_2_5_order,]

BOTTOM_ASSAULT <- as.data.frame(sub_by_assault_2_5[sub_by_assault_2_5_order,])

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

# barplot Cities with ASSAULT At or Below 2.5th Percentile in ascending order

BOTTOM_ASSAULT_BAR <- ggplot(BOTTOM_ASSAULT , aes(x = reorder(CITY,ASSAULT),
                                                  y = ASSAULT)) +
  geom_bar(fill = "#ca3886", stat = "identity") +
  geom_text(aes(label = ASSAULT), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Assault Lower 2.5% of Sample") +
  xlab("METRO") + ylab("ASSAULT per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

####################
```

```
# Assault barplot with lower 2.5% and outliers

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

ASSAULT_BAR <- ggarrange(BOTTOM_ASSAULT_BAR,
                         OUTLIER_ASSAULT_BAR,
                         ncol = 2,
                         nrow = 1)
ASSAULT_BAR

#########################################################
#########################################################
```

```
# US Crime 2018 features

# univariate distribution analysis

# Burglary

# density, histogram, boxplot, outliers, lower 2.5%
# (percentile) crime, qqplots, and shapiro-wilk tests

####################

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

#install.packages("gridExtra", dependencies = TRUE)
#library("gridExtra")

#install.packages("goft", dependencies = TRUE)
#library("goft")

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")

#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

####################
```

```
# Burglary

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

# Density

BURGLARY_DENSITY <- densityplot(~BURGLARY, data = US_CRIME_2018,
                                main="Burglary Density Plot",
                                col = "#2b507c")

# Histogram

BURGLARY_HISTOGRAM <- histogram(x=~BURGLARY,data=US_CRIME_2018,
                                type="density",
                                main="Burglary Histogram",
                                col = "#2b507c",
                                nint = 50)

# test distribution is normal

shapiro.test(x=US_CRIME_2018$BURGLARY)
```

```
# Normal QQ-Plot

BURGLARY_QQ_QNORM <- qqmath(x=~BURGLARY, data = US_CRIME_2018,
                            distribution = qnorm,
                            prepanel = prepanel.qqmathline,
                            panel = function(x, ...) {
                              panel.qqmathline(x, ...)
                              panel.qqmath(x, ...)
                            },
                            main = "Burglary Normal QQ-Plot \n SW-Test p-value = 0",
                            col = "#2b507c")

# test distribution is log normal

#install.packages("goft", dependencies = TRUE)
#library("goft")

lnorm_test(x=US_CRIME_2018$BURGLARY)

# Log-Normal QQ-Plot

BURGLARY_QQ_QLNORM <- qqmath(x=~BURGLARY, data = US_CRIME_2018,
                             distribution = qlnorm,
                             prepanel = prepanel.qqmathline,
                             panel = function(x, ...) {
                               panel.qqmathline(x, ...)
                               panel.qqmath(x, ...)
                             },
                             main = "Burglary Log-Normal QQ-Plot \n SW-Test p-value = 0.4335",
                             col = "#2b507c")
```

```
# combine 4 plots burglary

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(BURGLARY_DENSITY,
             BURGLARY_HISTOGRAM,
             BURGLARY_QQ_QNORM,
             BURGLARY_QQ_QLNORM,
             ncol=2)

##############################################

# BURGLARY

# find outliers based on boxplot

OutVals_burglary <- boxplot(US_CRIME_2018$BURGLARY)$out

which(US_CRIME_2018$BURGLARY %in% OutVals_burglary)

US_CRIME_2018[c(6,  81, 120, 128, 142, 155, 169, 196, 203, 227, 263),]

sub_burglary_outlier <- as.data.frame(US_CRIME_2018[c(6,  81, 120, 128,
                                                      142, 155, 169, 196,
                                                      203, 227, 263),])

sub_burglary_outlier_order <- order(sub_burglary_outlier$BURGLARY,decreasing = TRUE)

sub_burglary_outlier[sub_burglary_outlier_order,]

OUTLIER_BURGLARY <- as.data.frame(sub_burglary_outlier[sub_burglary_outlier_order,])
```

```
# barplot of outliers for Burglary

#install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

OUTLIER_BURGLARY_BAR <- ggplot(OUTLIER_BURGLARY, aes(x = reorder(CITY,-BURGLARY),
                                                      y = BURGLARY)) +
  geom_bar(fill = "#2b507c", stat = "identity") +
  geom_text(aes(label = BURGLARY), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Burglary Outliers of Sample") +
  xlab("METRO") + ylab("BURGLARY per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

####################

# Burglary 2.5th percentile

quantile(US_CRIME_2018$BURGLARY, .025)

####################
```

```
# at or below 2.5th percentile metro statistical areas in terms of burglary per 100,000

US_CRIME_2018[US_CRIME_2018$BURGLARY <= 137.430,]

sub_by_burglary_2_5 <- as.data.frame(US_CRIME_2018[US_CRIME_2018$BURGLARY <= 137.430,])

sub_by_burglary_2_5_order <- order(sub_by_burglary_2_5$BURGLARY,decreasing = FALSE)

sub_by_burglary_2_5[sub_by_burglary_2_5_order,]

BOTTOM_BURGLARY <- as.data.frame(sub_by_burglary_2_5[sub_by_burglary_2_5_order,])

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

# barplot Cities with BURGLARY At or Below 2.5th Percentile in ascending order

BOTTOM_BURGLARY_BAR <- ggplot(BOTTOM_BURGLARY , aes(x = reorder(CITY,BURGLARY),
                                                     y = BURGLARY)) +
  geom_bar(fill = "#2b507c", stat = "identity") +
  geom_text(aes(label = BURGLARY), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Burglary Lower 2.5% of Sample") +
  xlab("METRO") + ylab("BURGLARY per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

####################
```

```
# Burglary barplot with lower 2.5% and outliers

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

BURGLARY_BAR <- ggarrange(BOTTOM_BURGLARY_BAR,
                          OUTLIER_BURGLARY_BAR,
                          ncol = 2,
                          nrow = 1)
BURGLARY_BAR

###########################################################
###########################################################
```

```
# US Crime 2018 features

# univariate distribution analysis

# Larceny

# density, histogram, boxplot, outliers, lower 2.5%
# (percentile) crime, qqplots, and shapiro-wilk tests

####################

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

#install.packages("gridExtra", dependencies = TRUE)
#library("gridExtra")

#install.packages("goft", dependencies = TRUE)
#library("goft")

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")

#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

####################
```

```
# Larceny

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

# Density

LARCENY_DENSITY <- densityplot(~LARCENY, data = US_CRIME_2018,
                               main="Larceny Density Plot",
                               col = "#ff743b")

# Histogram

LARCENY_HISTOGRAM <- histogram(x=~LARCENY,data=US_CRIME_2018,
                               type="density",
                               main="Larceny Histogram",
                               col = "#ff743b",
                               nint = 50)

# test distribution is normal

shapiro.test(x=US_CRIME_2018$LARCENY)
```

```
# Normal QQ-Plot

LARCENY_QQ_QNORM <- qqmath(x=~LARCENY, data = US_CRIME_2018,
                           distribution = qnorm,
                           prepanel = prepanel.qqmathline,
                           panel = function(x, ...) {
                             panel.qqmathline(x, ...)
                             panel.qqmath(x, ...)
                           },
                           main = "Larceny Normal QQ-Plot \n SW-Test p-value = 0",
                           col = "#ff743b")

# test distribution is log normal

#install.packages("goft", dependencies = TRUE)
#library("goft")

lnorm_test(x=US_CRIME_2018$LARCENY)

# Log-Normal QQ-Plot

LARCENY_QQ_QLNORM <- qqmath(x=~LARCENY, data = US_CRIME_2018,
                            distribution = qlnorm,
                            prepanel = prepanel.qqmathline,
                            panel = function(x, ...) {
                              panel.qqmathline(x, ...)
                              panel.qqmath(x, ...)
                            },
                            main = "Larceny Log-Normal QQ-Plot \n SW-Test p-value = 0.04679",
                            col = "#ff743b")
```

```
# combine 4 plots larceny

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(LARCENY_DENSITY,
             LARCENY_HISTOGRAM,
             LARCENY_QQ_QNORM,
             LARCENY_QQ_QLNORM,
             ncol=2)

##################################################
```

```
# LARCENY

# find outliers based on boxplot

OutVals_larceny <- boxplot(US_CRIME_2018$LARCENY)$out

which(US_CRIME_2018$LARCENY %in% OutVals_larceny)

US_CRIME_2018[c(10,  93,  96, 169, 203, 209, 276, 286, 292),]

sub_larceny_outlier <- as.data.frame(US_CRIME_2018[c(10,  93,
                                                      96, 169, 203, 209,
                                                      276, 286, 292),])

sub_larceny_outlier_order <- order(sub_larceny_outlier$LARCENY,decreasing = TRUE)

sub_larceny_outlier[sub_larceny_outlier_order,]

OUTLIER_LARCENY <- as.data.frame(sub_larceny_outlier[sub_larceny_outlier_order,])
```

```
# barplot of outliers for Larceny

#install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

OUTLIER_LARCENY_BAR <- ggplot(OUTLIER_LARCENY, aes(x = reorder(CITY,-LARCENY),
                                                    y = LARCENY)) +
  geom_bar(fill = "#ff743b", stat = "identity") +
  geom_text(aes(label = LARCENY), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Larceny Outliers of Sample") +
  xlab("METRO") + ylab("LARCENY per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

####################

# Larceny 2.5th percentile

quantile(US_CRIME_2018$LARCENY, .025)

####################
```

```r
# at or below 2.5th percentile metro statistical areas in terms of larceny per 100,000

US_CRIME_2018[US_CRIME_2018$LARCENY <= 806.320,]

sub_by_larceny_2_5 <- as.data.frame(US_CRIME_2018[US_CRIME_2018$LARCENY <= 806.320,])

sub_by_larceny_2_5_order <- order(sub_by_larceny_2_5$LARCENY,decreasing = FALSE)

sub_by_larceny_2_5[sub_by_larceny_2_5_order,]

BOTTOM_LARCENY <- as.data.frame(sub_by_larceny_2_5[sub_by_larceny_2_5_order,])

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

# barplot Cities with LARCENY At or Below 2.5th Percentile in ascending order

BOTTOM_LARCENY_BAR <- ggplot(BOTTOM_LARCENY , aes(x = reorder(CITY,LARCENY),
                                                  y = LARCENY)) +
  geom_bar(fill = "#ff743b", stat = "identity") +
  geom_text(aes(label = LARCENY), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Larceny Lower 2.5% of Sample") +
  xlab("METRO") + ylab("LARCENY per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

####################
```

```r
# Larceny barplot with lower 2.5% and outliers

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

LARCENY_BAR <- ggarrange(BOTTOM_LARCENY_BAR,
                         OUTLIER_LARCENY_BAR,
                         ncol = 2,
                         nrow = 1)
LARCENY_BAR

###########################################################
###########################################################
```

```
# US Crime 2018 features

# univariate distribution analysis

# Vehicle

# density, histogram, boxplot, outliers, lower 2.5%
# (percentile) crime, qqplots, and shapiro-wilk tests

####################

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

#install.packages("gridExtra", dependencies = TRUE)
#library("gridExtra")

#install.packages("goft", dependencies = TRUE)
#library("goft")

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")

#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

####################
```

```
# Vehicle

#install.packages("lattice", dependencies = TRUE)
#library("lattice")

# Density

VEHICLE_DENSITY <- densityplot(~VEHICLE, data = US_CRIME_2018,
                               main="Vehicle Density Plot",
                               col = "#c82300")

# Histogram

VEHICLE_HISTOGRAM <- histogram(x=~VEHICLE,data=US_CRIME_2018,
                               type="density",
                               main="Vehicle Histogram",
                               col = "#c82300",
                               nint = 50)

# test distribution is normal

shapiro.test(x=US_CRIME_2018$VEHICLE)
```

```r
# Normal QQ-Plot

VEHICLE_QQ_QNORM <- qqmath(x=~VEHICLE, data = US_CRIME_2018,
                           distribution = qnorm,
                           prepanel = prepanel.qqmathline,
                           panel = function(x, ...) {
                             panel.qqmathline(x, ...)
                             panel.qqmath(x, ...)
                           },
                           main = "Vehicle Normal QQ-Plot \n SW-Test p-value = 0",
                           col = "#c82300")

# test distribution is log normal

#install.packages("goft", dependencies = TRUE)
#library("goft")

lnorm_test(x=US_CRIME_2018$VEHICLE)

# Log-Normal QQ-Plot

VEHICLE_QQ_QLNORM <- qqmath(x=~VEHICLE, data = US_CRIME_2018,
                            distribution = qlnorm,
                            prepanel = prepanel.qqmathline,
                            panel = function(x, ...) {
                              panel.qqmathline(x, ...)
                              panel.qqmath(x, ...)
                            },
                            main = "Vehicle Log-Normal QQ-Plot \n SW-Test p-value = 0.03367",
                            col = "#c82300")
```

```r
# combine 4 plots vehicle

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(VEHICLE_DENSITY,
             VEHICLE_HISTOGRAM,
             VEHICLE_QQ_QNORM,
             VEHICLE_QQ_QLNORM,
             ncol=2)

################################################
```

```r
# VEHICLE

# find outliers based on boxplot

OutVals_vehicle <- boxplot(US_CRIME_2018$VEHICLE)$out

which(US_CRIME_2018$VEHICLE %in% OutVals_vehicle)

US_CRIME_2018[c(5,  10,  19,  81, 169, 203, 209, 238, 286),]

sub_vehicle_outlier <- as.data.frame(US_CRIME_2018[c(5,  10,  19,  81, 169,
                                                     203, 209, 238, 286),])

sub_vehicle_outlier_order <- order(sub_vehicle_outlier$VEHICLE,decreasing = TRUE)

sub_vehicle_outlier[sub_vehicle_outlier_order,]

OUTLIER_VEHICLE <- as.data.frame(sub_vehicle_outlier[sub_vehicle_outlier_order,])
```

```
# barplot of outliers for Vehicle

#install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

OUTLIER_VEHICLE_BAR <- ggplot(OUTLIER_VEHICLE, aes(x = reorder(CITY,-VEHICLE),
                                                   y = VEHICLE)) +
  geom_bar(fill = "#c82300", stat = "identity") +
  geom_text(aes(label = VEHICLE), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Vehicle Outliers of Sample") +
  xlab("METRO") + ylab("VEHICLE per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

####################

# Vehicle 2.5th percentile

quantile(US_CRIME_2018$VEHICLE, .025)

####################
```

```
# at or below 2.5th percentile metro statistical areas in terms of vehicle per 100,000

US_CRIME_2018[US_CRIME_2018$VEHICLE <= 36.430,]

sub_by_vehicle_2_5 <- as.data.frame(US_CRIME_2018[US_CRIME_2018$VEHICLE <= 36.430,])

sub_by_vehicle_2_5_order <- order(sub_by_larceny_2_5$VEHICLE,decreasing = FALSE)

sub_by_vehicle_2_5[sub_by_vehicle_2_5_order,]

BOTTOM_VEHICLE <- as.data.frame(sub_by_vehicle_2_5[sub_by_larceny_2_5_order,])

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

# barplot Cities with VEHICLE At or Below 2.5th Percentile in ascending order

BOTTOM_VEHICLE_BAR <- ggplot(BOTTOM_VEHICLE , aes(x = reorder(CITY,VEHICLE),
                                                  y = VEHICLE)) +
  geom_bar(fill = "#c82300", stat = "identity") +
  geom_text(aes(label = VEHICLE), vjust = -0.3) +
  theme_pubclean() +
  ggtitle("Vehicle Lower 2.5% of Sample") +
  xlab("METRO") + ylab("VEHICLE per 100,000") +
  ggpubr::rotate_x_text() +
  theme(plot.title = element_text(hjust = 0.5))

####################
```

```r
# Vehicle barplot with lower 2.5% and outliers

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

VEHICLE_BAR <- ggarrange(BOTTOM_VEHICLE_BAR,
                         OUTLIER_VEHICLE_BAR,
                         ncol = 2,
                         nrow = 1)
VEHICLE_BAR

###########################################################
###########################################################
```

```r
###########################################################
###########################################################

# US Crime 2018 features

# correlation matrix

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
library("GGally")

######################

# correlation matrix

# Custom options.

ggcorr(
  CRIME_2018_FEAT,
  label = TRUE,
  name = "Sample Correlation",
  geom = "circle",
  max_size = 20,
  min_size = 4,
  size = 4,
  hjust = 0.75,
  nbreaks = 6,
  angle = 0,
  palette = "PuOr")
```

```
############################################################
############################################################

# US Crime 2018 features

# contour plots and ggpairs 2d combination plots

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

# install.packages("scales")
library("scales")

######################
```

```
# ggpairs matrix

# upper =  scatterplot

# diagnol = density

# lower = contour plot

upperfun <- function(data,mapping){
  ggplot(data = data, mapping = mapping)+
    geom_point(alpha = .2, col="#0D0887FF")
}

diagfun <- function(data,mapping){
  ggplot(data = data, mapping = mapping)+
    geom_density(fill = "#B12A90FF", colour = "#0D0887FF")
}

lowerfun <- function(data,mapping){
  ggplot(data = data, mapping = mapping)+
    stat_density2d(aes(fill = stat(level)), geom="polygon") +
    scale_fill_viridis_c(option = "plasma") +
    theme(legend.position = "magma")
}
ggpairs(CRIME_2018_FEAT,upper = list(continuous = wrap(upperfun)),
        lower = list(continuous = wrap(lowerfun)),
        diag = list(continuous = wrap(diagfun))
)

############################################################
############################################################
```

```
# US Crime 2018 original variables

# multivariate crime analysis

# Test multivariate normality and chi-square plot

######################

# install.packages("goft", dependencies = TRUE)
# library("goft")

# install.packages("car", dependencies = TRUE)
library("car")

######################

# https://ggplot2.tidyverse.org/reference/geom_qq.html

######################

# Shapiro-Wilk test for multivariate normality

# A generalization of Shapiro-Wilk test for multivariate normality
# (Villasenor-Alva and GonzalezEstrada, 2009).

#install.packages("goft", dependencies = TRUE)
# library("goft")

mvshapiro_test(X=data.matrix(CRIME_2018_FEAT))
```

```
##########################################################
##########################################################
##########################################################
##########################################################

# US Crime 2018 standardized features

# univariate distribution analysis

# descriptives

###################

#install.packages("psych", dependencies = TRUE)
#library("psych")

describe(SCALED_CRIME_2018_FEAT)

summary(SCALED_CRIME_2018_FEAT)

############################################################
############################################################
```

```
##############################################################
##############################################################

# US Crime 2018 standardized features

# bivariate distribution analysis

# scatter matrix

##############################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

#############################

# https://ggobi.github.io/ggally/rd.html#ggmatrix

#############################

CRIME_scatM_S_FEAT <- ggpairs(SCALED_CRIME_2018_FEAT,
                              lower = list(continuous = wrap("points", alpha = 0.25)))
CRIME_scatM_S_FEAT

##############################################################
##############################################################
##############################################################
##############################################################
```

```
# Crime 2018 standardized features

# PCA

######################

# eigenvalues of correlation matrix

eigen(x=cor(SCALED_CRIME_2018_FEAT))

# PCA of standardized features

SCALED_CRIME_2018_FEAT_PCA <- prcomp(CRIME_2018_FEAT, scale = TRUE)
SCALED_CRIME_2018_FEAT_PCA

summary(SCALED_CRIME_2018_FEAT_PCA)

SCALED_CRIME_2018_FEAT_PCA$sdev

SCALED_CRIME_2018_FEAT_PCA$rotation

SCALED_CRIME_2018_FEAT_PCA$center

SCALED_CRIME_2018_FEAT_PCA$scale

SCALED_CRIME_2018_FEAT_PCA$x

######################################################
######################################################
```

```
# Crime 2018 standardized features

# create dataframe with PCA scores

########################

# 7-d pca scores

PCA_CRIME_2018 <- as.data.frame(SCALED_CRIME_2018_FEAT_PCA$x)
PCA_CRIME_2018

# change PCAi to Yi notation

y1 <- PCA_CRIME_2018$PC1
y2 <- PCA_CRIME_2018$PC2
y3 <- PCA_CRIME_2018$PC3
y4 <- PCA_CRIME_2018$PC4
y5 <- PCA_CRIME_2018$PC5
y6 <- PCA_CRIME_2018$PC6
y7 <- PCA_CRIME_2018$PC7

PCA_CRIME_2018 <- data.frame(y1, y2, y3, y4, y5, y6, y7)
PCA_CRIME_2018

# rename CITY for final dataframe

CITY <- CRIME_2018$CITY

# dataframe of pca scores with CITY variable

PCA_CRIME_2018 <- as.data.frame(cbind(CITY, PCA_CRIME_2018))
PCA_CRIME_2018
```

```
# set rownames (again)

row.names(PCA_CRIME_2018) <- PCA_CRIME_2018$CITY

# check rownames

PCA_CRIME_2018

# standardized pca scores -- PC1, PC2, PC3

PC1_PC2_PC3_CRIME <- as.data.frame(PCA_CRIME_2018[,2:4])
PC1_PC2_PC3_CRIME

# set rownames (again)

row.names(PC1_PC2_PC3_CRIME) <- PCA_CRIME_2018$CITY

# check rownames

PC1_PC2_PC3_CRIME

# see ranges of pc1, pc2, pc3

summary(PC1_PC2_PC3_CRIME)

##################################################################
##################################################################
```

```
########################################################################
########################################################################

# Crime 2018 standardized features

# percent of explained variance pca of scaled 2018 crime data

###############

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")
#install.packages("factoextra", dependencies = TRUE)
library("factoextra")

###############

fviz_screeplot(SCALED_CRIME_2018_FEAT_PCA, addlabels = TRUE, ylim = c(0, 90)) +
  labs(title = "Explained Standardized Sample Variance by Principal Component",
       subtitle = " for US Crime 2018",
       x = "Standardized Sample Principal Components",
       y = "% of Explained Standardized Sample Variance") +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))

########################################################################
########################################################################
```

```
# Crime 2018 standardized features

# Variable contributions to the principal axes scaled 2018 crime data

#######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("factoextra", dependencies = TRUE)
#library("factoextra")

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")

#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

#######################
```

```
# Percent Contributions of standardized variables to PC1

CONT_PC1 <- fviz_contrib(SCALED_CRIME_2018_FEAT_PCA,
                         choice = "var",
                         axes = 1,
                         top = 7,
                         title = "Contributions to y1") +
  theme(plot.title = element_text(hjust = 0.5))

# Percent Contributions of standardized variables to PC2

CONT_PC2 <- fviz_contrib(SCALED_CRIME_2018_FEAT_PCA,
                         choice = "var",
                         axes = 2,
                         top = 7,
                         title = "Contributions to y2") +
  theme(plot.title = element_text(hjust = 0.5))

# Percent Contributions of standardized variables to PC3

CONT_PC3 <- fviz_contrib(SCALED_CRIME_2018_FEAT_PCA,
                         choice = "var",
                         axes = 3,
                         top = 7,
                         title = "Contributions to y3") +
  theme(plot.title = element_text(hjust = 0.5))


####################
```

```
# grid of plots contributions to principal components of
# standardized variables Crime 2018

#install.packages("magrittr", dependencies = TRUE)
#library("magrittr")
#install.packages("ggpubr", dependencies = TRUE)
#library("ggpubr")

CONT_PCS_PLOT_1_3 <- ggarrange(CONT_PC1,
                               CONT_PC2,
                               CONT_PC3,
                               ncol = 2,
                               nrow = 2)
CONT_PCS_PLOT_1_3

######################################################################
######################################################################
```

```
# Crime 2018 PCA from standardized features

# correlation pc's with original components

##############################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

##############################

# combined 2018 standardized features and pca scores

SCALED_PCA_AND_SCALED_CRIME_2018 <- as.data.frame(cbind(PCA_CRIME_2018,
                                                        SCALED_CRIME_2018_FEAT))

##############################
```

```
# Custom options angle = 0
ggcorr(
  SCALED_PCA_AND_SCALED_CRIME_2018[,-1],
  label = TRUE,
  name = "Sample Correlation",
  geom = "circle",
  max_size = 20,
  min_size = 4,
  size = 4,
  hjust = 0.75,
  nbreaks = 6,
  angle = 0,
  palette = "PuOr")

#################################################################
#################################################################
```

```
# Crime 2018 PCA from standardized features

# scatterplot matrix of PC1, PC2, PC3 components

# Scatterplot of pairs PC1, PC2, PC3

##############################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

##############################
```

```r
# Scatterplot of pairs PC1, PC2, PC3

# Scatterplot PC1, PC2

PC1_PC2_SCATTER <- ggplot(PCA_CRIME_2018, aes(x=y1, y=y2)) +
  geom_text(label=rownames(PCA_CRIME_2018)) +
  #    ggtitle("Scatterplot of y2 ~ y1") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)") +
  theme(plot.title = element_text(hjust = 0.5))

####################

# Scatterplot PC1, PC3

PC1_PC3_SCATTER <- ggplot(PCA_CRIME_2018, aes(x=y1, y=y3)) +
  geom_text(label=rownames(PCA_CRIME_2018)) +
  #      ggtitle("Scatterplot of y3 ~ y1") +
  labs(x = "y1 (63.05%)", y = "y3 (9.383%)") +
  theme(plot.title = element_text(hjust = 0.5))

####################

# Scatterplot PC2, PC3

PC2_PC3_SCATTER <- ggplot(PCA_CRIME_2018, aes(x=y2, y=y3)) +
  geom_text(label=rownames(PCA_CRIME_2018)) +
  ggtitle("Scatterplot of y3 ~ y2") +
  labs(x = "y2 (10.99%)", y = "y3 (9.383%)") +
  theme(plot.title = element_text(hjust = 0.5))
```

```r
####################################################################
####################################################################
####################################################################
####################################################################

# Clustering Methods

# Partitioning

# K-means

####################################################################
####################################################################
```

```r
# K-means

# Scaled Crime 2018 Features

# PC1,PC2, PC3 from Scaled Crime 2018 Features

# Estimating the optimal number of clusters

# NbClust Method

#######################
```

```r
# install.packages("factoextra")
# library("factoextra")

# install.packages("NbClust", dependencies = TRUE)
library("NbClust")

# install.packages("magrittr", dependencies = TRUE)
# library("magrittr")

# install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

#####################
```

```r
# Scaled Crime 2018 Features

# install.packages("NbClust", dependencies = TRUE)
# library("NbClust")

NBCLUST_SCALE_CRIME_2018_FEAT_KMEANS <- NbClust(SCALED_CRIME_2018_FEAT,
                                                distance = "euclidean",
                                                min.nc = 2,
                                                max.nc = 6,
                                                method = "kmeans")

# install.packages("factoextra", dependencies = TRUE)
# library("factoextra")

NBCLUST_SCALE_CRIME_2018_FEAT_KMEANS_BOX <- fviz_nbclust(NBCLUST_SCALE_CRIME_2018_FEAT_KMEANS,
                                                         barfill = "steelblue",
                                                         barcolor = "steelblue") +
  labs(title= "NbClust, Black-Box Method, k-Means, Input Standardized Crime 2018") +
  xlab("# of clusters k") +
  ylab("Freq. Among Indices") +
  theme(plot.title = element_text(hjust = 0.5))

###########################################################
```

```r
# PC1,PC2, PC3 from Scaled Crime 2018 Features

# install.packages("NbClust", dependencies = TRUE)
# library("NbClust")

NBCLUST_PC1_PC2_PC3_CRIME_2018_KMEANS <- NbClust(PC1_PC2_PC3_CRIME,
                                                 distance = "euclidean",
                                                 min.nc = 2,
                                                 max.nc = 6,
                                                 method = "kmeans")

# install.packages("factoextra", dependencies = TRUE)
# library("factoextra")

NBCLUST_PC1_PC2_PC3_CRIME_2018_KMEANS_BOX <- fviz_nbclust(NBCLUST_PC1_PC2_PC3_CRIME_2018_KMEANS
                                                          barfill = "steelblue",
                                                          barcolor = "steelblue") +
  labs(title= "NbClust, Black-Box Method, k-Means, Input y1, y2, y3") +
  xlab("# of clusters k") +
  ylab("Freq. Among Indices") +
  theme(plot.title = element_text(hjust = 0.5))

########################################################
```

```
# grid of plots for OPTIMAL K

# install.packages("magrittr", dependencies = TRUE)
# library("magrittr")
# install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

OPTIMAL_K_KMEAN_CRIME_2018 <- ggarrange(NBCLUST_SCALE_CRIME_2018_FEAT_KMEANS_BOX,
                                        NBCLUST_PC1_PC2_PC3_CRIME_2018_KMEANS_BOX,
                                        ncol = 1,
                                        nrow = 2)

####################################################################
####################################################################
####################################################################

# K-means using eclust() in "factoextra" package

# k = 3

# Scaled Crime 2018 Features

# PC1,PC2, PC3 from Scaled Crime 2018 Features

# create dataframe with  CITY, 7 Crime 2018 variables,
# PC1 PC2 PC3, k=3 kmean cluster assignment(s), and rownames

######################
```

```
# install.packages("factoextra")
# library("factoextra")

#install.packages("plyr", dependencies = TRUE)
library("plyr")

# Compute k-means with k = 3 with scaled crime 2018 features

# install.packages("factoextra", dependencies = TRUE)
# library("factoextra")

set.seed(125)
KM3_SCALED_CRIME <- eclust(SCALED_CRIME_2018_FEAT, "kmeans",
                           k = 3, nstart = 25, graph = FALSE)
KM3_SCALED_CRIME$size

# Compute k-means with k = 3 PC 1, 2, 3

# install.packages("factoextra", dependencies = TRUE)
# library("factoextra")

set.seed(126)
KM3_PC1_PC2_PC3_CRIME <- eclust(PC1_PC2_PC3_CRIME, "kmeans",
                                k = 3, nstart = 25, graph = FALSE)
KM3_PC1_PC2_PC3_CRIME$size

##############################
```

```
# create factors from cluster assignments

# CITY <- CRIME_2018$CITY
KM3_SCALED_ASSIGN <- as.character(KM3_SCALED_CRIME$cluster)
KM3_PC1_PC2_PC3_ASSIGN <- as.character(KM3_PC1_PC2_PC3_CRIME$cluster)

#install.packages("plyr", dependencies = TRUE)
#library("plyr")

KM3_SCALED_ASSIGN <- revalue(KM3_SCALED_ASSIGN, c("1"="2", "2"="3", "3"="1"))
KM3_PC1_PC2_PC3_ASSIGN <- revalue(KM3_PC1_PC2_PC3_ASSIGN, c("1"="2", "2"="1", "3"="3"))

# create dataframe to combine all the data and results

CRIME_2018_ASSIGN <- as.data.frame(cbind(
  CITY,
  CRIME_2018_FEAT,
  PC1_PC2_PC3_CRIME,
  KM3_SCALED_ASSIGN,
  KM3_PC1_PC2_PC3_ASSIGN)
)

CRIME_2018_ASSIGN

# set rownames CITY_CRIME_2018

row.names(CRIME_2018_ASSIGN) <- CRIME_2018_ASSIGN$CITY

# summary

summary(CRIME_2018_ASSIGN)
```

```
####################################################################
####################################################################
####################################################################

# K-means using eclust() in "factoextra" package

# k = 3

# Scaled Crime 2018 Features

# PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# Cluster City Names, Cluster Mean Vectors, Cluster sd Vectors,

# check for differences in cluster assignments

#####################
```

```
#####################

# cluster city names

# k-mean 3 cluster solution on standardized features

# cluster 1

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$KM3_SCALED_ASSIGN == "1",1])

# cluster 2

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$KM3_SCALED_ASSIGN == "2",1])

# cluster 3

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$KM3_SCALED_ASSIGN == "3",1])

####################
```

```
# cluster city names

# k-mean 3 cluster solution on PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# cluster 1

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$KM3_PC1_PC2_PC3_ASSIGN == "1",1])

# cluster 2

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$KM3_PC1_PC2_PC3_ASSIGN == "2",1])

# cluster 3

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$KM3_PC1_PC2_PC3_ASSIGN == "3",1])

####################

# 1-way table of cluster assignments

table(KM3_SCALED_ASSIGN)

table(KM3_PC1_PC2_PC3_ASSIGN)

#####################

# 2-way table of cluster assignments (confusion matrix)

table(KM3_SCALED_ASSIGN, KM3_PC1_PC2_PC3_ASSIGN)
```

```
####################

# city names for misclassifications

# (in cluster "1" for KM3_SCALED_ASSIGN) and (in cluster "2" for KM3_PC1_PC2_PC3_ASSIGN )

CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$KM3_SCALED_ASSIGN == "1"
                  &
                  CRIME_2018_ASSIGN$KM3_PC1_PC2_PC3_ASSIGN == "2",
                  c(12,13)]

# (in cluster "2" for KM3_SCALED_ASSIGN) and (in cluster "3" for KM3_PC1_PC2_PC3_ASSIGN )

CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$KM3_SCALED_ASSIGN == "2"
                  &
                  CRIME_2018_ASSIGN$KM3_PC1_PC2_PC3_ASSIGN == "3",
                  c(12,13)]

##########################

# cluster means, k-means, k=3, standardized Crime 2018 Features

aggregate(CRIME_2018_ASSIGN[, 2:8], list(CRIME_2018_ASSIGN$KM3_SCALED_ASSIGN), mean)

##########################

# cluster means, k-means, k=3, PC1, PC2, PC3 derived from Scaled Crime 2018 Features

aggregate(CRIME_2018_ASSIGN[, 2:8], list(CRIME_2018_ASSIGN$KM3_PC1_PC2_PC3_ASSIGN), mean)

#########################
```

```
#install.packages("psych", dependencies = TRUE)
# library("psych")

# original sample means

describe(CRIME_2018_FEAT)

##############################################################
##############################################################
##############################################################
```

```
# K-means using eclust() in "factoextra" package

# k = 3

# Scaled Crime 2018 Features

# Scatterplots on PC1, PC2, PC3 of cluster assignments

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

#####################
```

```r
# 3 group k-mean scatterplot with CITY labels input standardized crime 2018 data.
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

KM3_SCALE_CRIME_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN,
                                    aes(x=y1, y=y2, color=KM3_SCALED_ASSIGN)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  # ggtitle("k-Means, k=3, Input Standardized Crime 2018") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF","#009788")) +
  # theme(plot.title = element_text(hjust = 0.5)) +
  #  theme(legend.position = "none")
  theme(legend.position = "bottom")

####################
```

```r
# 3 group k-mean scatterplot with CITY labels input standardized crime 2018 data.
# Plotted on PC1, PC3

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

KM3_SCALE_CRIME_DIM1_DIM3 <- ggplot(CRIME_2018_ASSIGN,
                                    aes(x=y1, y=y3, color=KM3_SCALED_ASSIGN)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  #  ggtitle("k-Means, k=3, Input Standardized Crime 2018") +
  labs(x = "y1 (63.05%)", y = "y3 (9.383%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  #  theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "none")

####################
```

```r
# change legend position for final plot

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

get_legend<-function(myggplot){
  tmp <- ggplot_gtable(ggplot_build(myggplot))
  leg <- which(sapply(tmp$grobs, function(x) x$name) == "guide-box")
  legend <- tmp$grobs[[leg]]
  return(legend)
}

legend <- get_legend(KM3_SCALE_CRIME_DIM1_DIM2)

KM3_SCALE_CRIME_DIM1_DIM2 <- KM3_SCALE_CRIME_DIM1_DIM2 +
                             theme(legend.position = "none")

####################
```

```
# Combine 2 scatterplots, K-Means, K=3

# Input Scaled Crime 2018 Features

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(KM3_SCALE_CRIME_DIM1_DIM2,
             KM3_SCALE_CRIME_DIM1_DIM3,
             legend,
             ncol=2,
             nrow= 2,
             layout_matrix = rbind(c(1,2), c(3,3)),
             widths = c(2.7, 2.7), heights = c(2.5, 0.2),
             top = text_grob("k-Means, k=3, Input Standardized Crime 2018,
                             Plotted on y1, y2, y3",
                             color = "black", face = "bold", size = 14))

##########################################################################
##########################################################################
```

```
# K-means using eclust() in "factoextra" package

# k = 3

# PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# Scatterplots on PC1, PC2, PC3 of cluster assignments

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

#####################
```

```
# 3 group k-mean scatterplot with CITY labels, input PC1, PC2, PC3
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

KM3_PC1_PC2_PC3_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN,
                                    aes(x=y1, y=y2, color=KM3_PC1_PC2_PC3_ASSIGN)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  #  ggtitle("k-Means, k=3, Input y1, y2, y3") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  theme(plot.title = element_text(hjust = 0.5)) +
  #  theme(legend.position = "none")
  theme(legend.position = "bottom")

###################
```

```
# 3 group k-mean scatterplot with CITY labels, input PC1, PC2, PC3
# Plotted on PC1, PC3

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

KM3_PC1_PC2_PC3_DIM1_DIM3 <- ggplot(CRIME_2018_ASSIGN,
                                    aes(x=y1, y=y3, color=KM3_PC1_PC2_PC3_ASSIGN)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  #  ggtitle("k-Means, k=3, Input y1, y2, y3") +
  labs(x = "y1 (63.05%)", y = "y3 (9.383%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  # theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "none")

####################
```

```
# change legend position for final plot

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

get_legend<-function(myggplot){
  tmp <- ggplot_gtable(ggplot_build(myggplot))
  leg <- which(sapply(tmp$grobs, function(x) x$name) == "guide-box")
  legend <- tmp$grobs[[leg]]
  return(legend)
}

legend <- get_legend(KM3_PC1_PC2_PC3_DIM1_DIM2)

KM3_PC1_PC2_PC3_DIM1_DIM2 <- KM3_PC1_PC2_PC3_DIM1_DIM2 +
                             theme(legend.position = "none")

```

```
####################

# Combine 2 scatterplots, K-Means K=3

# Input PC1, PC2, PC3

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(KM3_PC1_PC2_PC3_DIM1_DIM2,
             KM3_PC1_PC2_PC3_DIM1_DIM3,
             legend,
             ncol=2,
             nrow=2,
             layout_matrix = rbind(c(1,2), c(3,3)),
             widths = c(2.7, 2.7), heights = c(2.5, 0.2),
             top = text_grob("k-Means, k=3, Input y1, y2, y3, Plotted on y1, y2, y3",
                             color = "black", face = "bold", size = 14))

####################################################################
####################################################################
####################################################################
```

```
# K-means using eclust() in "factoextra" package

# k = 3

# Standardized Crime 2018 Features

# Scattermatrix on original dimensions

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

##############################
# Scaled Crime 2018 Features

# ggpairs

# k = 3

# K-Means

p <- ggpairs(CRIME_2018_ASSIGN, c(2,3,4,5,6,7,8,12),
             mapping = ggplot2::aes_string(color = "KM3_SCALED_ASSIGN"))

for(i in 1:p$nrow) {
  for(j in 1:p$ncol){
    p[i,j] <- p[i,j] +
      scale_fill_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
      scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788"))
  }
}

p
```

```
#################################################################
#################################################################

# K-means using eclust() in "factoextra" package

# k = 3

# PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# Scattermatrix on original dimensions

#####################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

##############################
```

```r
# PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# ggpairs

# k = 3

# K-Means

p <- ggpairs(CRIME_2018_ASSIGN, c(2,3,4,5,6,7,8,13),
             mapping = ggplot2::aes_string(color = "KM3_PC1_PC2_PC3_ASSIGN"))

for(i in 1:p$nrow) {
  for(j in 1:p$ncol){
    p[i,j] <- p[i,j] +
      scale_fill_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
      scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788"))
  }
}

p

####################################################################
####################################################################
####################################################################
####################################################################
```

```r
# Clustering Methods

# Hierarchical

# Agglomerative

####################################################################
####################################################################

# Distance Matrices

# Scaled Crime 2018 Features

# PC1, PC2, PC3 from Scaled Crime 2018 Features

######################
```

```r
# Euclidean Distance Matrix on Scaled Crime 2018 Features

DIST_SCALED_CRIME <- dist(SCALED_CRIME_2018_FEAT, method = "euclidean")
DIST_SCALED_CRIME

# Subset the first 5 columns and rows on Scaled Crime 2018 Features

round(as.matrix(DIST_SCALED_CRIME)[1:5, 1:5], 1)

######################

# Euclidean Distance Matrix on PC1, PC2, PC3

DIST_PC1_PC2_PC3_CRIME <- dist(PC1_PC2_PC3_CRIME, method = "euclidean")
DIST_PC1_PC2_PC3_CRIME

# Subset the first 5 columns and rows on PC1, PC2, PC3

round(as.matrix(DIST_PC1_PC2_PC3_CRIME)[1:5, 1:5], 1)
```

```
####################################################################
####################################################################
####################################################################

# Agglomerative Clustering

# WARD.D2

# Scaled Crime 2018 Features

# PC1,PC2, PC3 from Scaled Crime 2018 Features

# Estimating the optimal number of clusters

# NbClust Method

######################

# install.packages("factoextra")
# library("factoextra")

# install.packages("NbClust", dependencies = TRUE)
# library("NbClust")

# install.packages("magrittr", dependencies = TRUE)
# library("magrittr")

# install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

#####################
```

```
# Scaled Crime 2018 Features

# install.packages("NbClust", dependencies = TRUE)
# library("NbClust")

NBCLUST_SCALE_CRIME_2018_FEAT_WARD.D2 <- NbClust(SCALED_CRIME_2018_FEAT,
                                                 distance = "euclidean",
                                                 min.nc = 2,
                                                 max.nc = 6,
                                                 method = "ward.D2")

# install.packages("factoextra", dependencies = TRUE)
# library("factoextra")

NBCLUST_SCALE_CRIME_2018_FEAT_WARD.D2_BOX <- fviz_nbclust(NBCLUST_SCALE_CRIME_2018_FEAT_WARD.D2
                                                          barfill = "steelblue",
                                                          barcolor = "steelblue") +
  labs(title= "NbClust, Black-Box Method, Ward, Input Standardized US Crime 2018") +
  xlab("# of clusters k") +
  ylab("Freq. Among Indices") +
  theme(plot.title = element_text(hjust = 0.5))

#########################################################
```

```
# PC1, PC2, PC3 from Scaled Crime 2018 Features

# install.packages("NbClust", dependencies = TRUE)
# library("NbClust")

NBCLUST_PC1_PC2_PC3_CRIME_2018_WARD.D2 <- NbClust(PC1_PC2_PC3_CRIME,
                                                  distance = "euclidean",
                                                  min.nc = 2,
                                                  max.nc = 6,
                                                  method = "ward.D2")

# install.packages("factoextra", dependencies = TRUE)
# library("factoextra")

NBCLUST_PC1_PC2_PC3_CRIME_2018_WARD.D2_BOX <- fviz_nbclust(NBCLUST_PC1_PC2_PC3_CRIME_2018_WARD.D2
                                                          barfill = "steelblue",
                                                          barcolor = "steelblue") +
  labs(title= "NbClust, Black-Box Method, Ward, Input y1, y2, y3") +
  xlab("# of clusters k") +
  ylab("Freq. Among Indices") +
  theme(plot.title = element_text(hjust = 0.5))

##########################################################
```

```
#########################################################

# grid of plots for OPTIMAL K

# install.packages("magrittr", dependencies = TRUE)
# library("magrittr")
# install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

OPTIMAL_WARD.D2_CRIME_2018 <- ggarrange(NBCLUST_SCALE_CRIME_2018_FEAT_WARD.D2_BOX,
                                        NBCLUST_PC1_PC2_PC3_CRIME_2018_WARD.D2_BOX,
                                        ncol = 1,
                                        nrow = 2)

####################################################################
####################################################################
####################################################################
```

```
# Agglomerative Clustering

#  WARD.D2

# Compute algorithm

# Scaled Crime 2018 Features

# PC1, PC2, PC3 from Scaled Crime 2018 Features

######################

# Compute "ward.d2" using Scaled Crime 2018 Features

set.seed(150)
SCALED_WARD.D2 <- hclust(d = DIST_SCALED_CRIME, method = "ward.D2")

# Compute "ward.d2" using PC1, PC2, PC3

set.seed(151)
PC1_PC2_PC3_WARD.D2 <- hclust(d = DIST_PC1_PC2_PC3_CRIME, method = "ward.D2")

#####################################################################
#####################################################################
#####################################################################
```

```
# Agglomerative Clustering

# WARD.D2

# K = 3

# add cluster assignments

# Scaled Crime 2018 Features

# PC1, PC2, PC3 from Scaled Crime 2018 Features

######################

#install.packages("plyr", dependencies = TRUE)
#library("plyr")

#####################

# Cut tree into 3 clusters "ward.d2" from Scaled Crime 2018 Features

K3_SCALED_WARD.D2 <- as.character(cutree(SCALED_WARD.D2, k = 3))

table(K3_SCALED_WARD.D2)

# Cut tree into 3 clusters "ward.d2" from PC1, PC2, PC3

K3_PC1_PC2_PC3_WARD.D2 <- as.character(cutree(PC1_PC2_PC3_WARD.D2, k = 3))

table(K3_PC1_PC2_PC3_WARD.D2)
```

```
#install.packages("plyr", dependencies = TRUE)
#library("plyr")

K3_SCALED_WARD.D2 <- revalue(K3_SCALED_WARD.D2, c("1"="3", "2"="2", "3"="1"))
K3_PC1_PC2_PC3_WARD.D2 <- revalue(K3_PC1_PC2_PC3_WARD.D2, c("1"="3", "2"="2", "3"="1"))

# create dataframe to combine all the data and results

CRIME_2018_ASSIGN <- as.data.frame(cbind(
  CRIME_2018_ASSIGN,
  K3_SCALED_WARD.D2,
  K3_PC1_PC2_PC3_WARD.D2)
)

CRIME_2018_ASSIGN

# set rownames CITY_CRIME_2018

row.names(CRIME_2018_ASSIGN) <- CRIME_2018_ASSIGN$CITY

# summary

summary(CRIME_2018_ASSIGN)

####################################################################
####################################################################
####################################################################
```

```
# WARD.D2

# k = 3

# Scaled Crime 2018 Features

# PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# Cluster City Names, Cluster Mean Vectors, Cluster sd Vectors,
# check for differences in cluster assignments

####################

# cluster city names

# WARD.D2 3 cluster solution on standardized features

# cluster 1

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_SCALED_WARD.D2 == "1",1])

# cluster 2

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_SCALED_WARD.D2 == "2",1])

# cluster 3

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_SCALED_WARD.D2 == "3",1])

####################
```

```
# cluster city names

# WARD.D2 3 cluster solution on PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# cluster 1

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_PC1_PC2_PC3_WARD.D2== "1",1])

# cluster 2

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_PC1_PC2_PC3_WARD.D2 == "2",1])

# cluster 3

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_PC1_PC2_PC3_WARD.D2 == "3",1])

####################

# 1-way table of cluster assignments

table(K3_SCALED_WARD.D2)

table(K3_PC1_PC2_PC3_WARD.D2)

####################

# 2-way table of cluster assignments (confusion matrix)

table(K3_SCALED_WARD.D2, K3_PC1_PC2_PC3_WARD.D2)

####################
```

```r
# city names for misclassifications

# (in cluster "1" for K3_SCALED_WARD.D2) and (in cluster "2" for K3_PC1_PC2_PC3_WARD.D2)

CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_SCALED_WARD.D2 == "1"
                  &
                  CRIME_2018_ASSIGN$K3_PC1_PC2_PC3_WARD.D2 == "2", c(14,15)]

# (in cluster "2" for K3_SCALED_WARD.D2) and (in cluster "3" for KM3_PC1_PC2_PC3_ASSIGN)

CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_SCALED_WARD.D2 == "2"
                  &
                  CRIME_2018_ASSIGN$K3_PC1_PC2_PC3_WARD.D2 == "3", c(14,15)]

# (in cluster "3" for K3_SCALED_WARD.D2) and (in cluster "2" for K3_PC1_PC2_PC3_WARD.D2)

CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_SCALED_WARD.D2 == "3"
                  &
                  CRIME_2018_ASSIGN$K3_PC1_PC2_PC3_WARD.D2 == "2", c(14,15)]

##########################

# cluster means, WARD.D2, k=3, standardized Crime 2018 Features

aggregate(CRIME_2018_ASSIGN[, 2:8], list(CRIME_2018_ASSIGN$K3_SCALED_WARD.D2), mean)

##########################
```

```r
#install.packages("psych", dependencies = TRUE)
# library("psych")

# original sample means

describe(CRIME_2018_FEAT)

######################################################################
######################################################################
######################################################################
```

```r
# Agglomerative Clustering

# WARD.D2

# K = 3

# Dendrograms

# Scaled Crime 2018 Features

######################

# install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")

# install.packages("factoextra")
# library("factoextra")

######################
```

```
# WARD.D2, Input Scaled Crime 2018, Rectangle Dendrogram, K=3

DEND_RECTANGLE_K3_SCALED_WARD.D2 <- fviz_dend(SCALED_WARD.D2, k = 3, # Cut in three groups
                                              cex = 0.5, # label size
                                              palette = c("#009788","#0080FF", "#FE3A07"),
                                              color_labels_by_k = TRUE, # color labels by groups
                                              rect = TRUE, # Add rectangle around groups
                                              # main = NULL,
                                              main = "Ward, k=3, Input Standardized Crime 2018,
                                              Rectangular Dendrogram",
                                              rect_border = c("#009788","#0080FF", "#FE3A07"),
                                              rect_fill = TRUE) +
                                              theme(plot.title = element_text(hjust = 0.5))

######################################################################
######################################################################
######################################################################

# Agglomerative Clustering

# WARD.D2

# K = 3

# Dendrograms

# PC1, PC2, PC3 from Scaled Crime 2018 Features

######################
```

```
# install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")

# install.packages("factoextra")
# library("factoextra")

######################

# WARD.D2, Input PC1, PC2, PC3, Rectangle Dendrogram, K=3

DEND_RECTANGLE_K3_PC1_PC2_PC3_WARD.D2 <- fviz_dend(PC1_PC2_PC3_WARD.D2, k = 3, # Cut in two groups
                                                   cex = 0.5, # label size
                                                   palette = c("#009788", "#FE3A07", "#0080FF"),
                                                   color_labels_by_k = TRUE, # color labels by groups
                                                   rect = TRUE, # Add rectangle around groups
                                                   main = "Ward, k=3, Input y1, y2, y3,
                                                   Rectangular Dendrogram",
                                                   rect_border = c("#009788", "#FE3A07", "#0080FF"),
                                                   rect_fill = TRUE) +
                                                   theme(plot.title = element_text(hjust = 0.5))




######################################################################
######################################################################
######################################################################
```

```
# Agglomerative Clustering

# WARD.D2

# K = 3

# Scaled Crime 2018 Features

# Scatterplots on PC1, PC2, PC3 of cluster assignments and ggpairs matrix

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

#####################
```

```
# 3 group WARD.D2 scatterplot with CITY labels input scaled crime 2018 data.
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_SCALED_WARD.D2_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN, aes(x=y1, y=y2,
                                                              color=K3_SCALED_WARD.D2)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  #  ggtitle("Wards, k=3, Input Standardized Crime 2018") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c( "#FE3A07", "#0080FF", "#009788")) +
  #  theme(plot.title = element_text(hjust = 0.5)) +
  #  theme(legend.position = "none")
  theme(legend.position = "bottom")

####################
```

```
# 3 group WARD.D2 scatterplot with CITY labels input scaled crime 2018 data.
# Plotted on PC1, PC3

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_SCALED_WARD.D2_DIM1_DIM3 <- ggplot(CRIME_2018_ASSIGN, aes(x=y1, y=y3,
                                                              color=K3_SCALED_WARD.D2)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  # ggtitle("Wards, k=3, Input Scaled Crime 2018") +
  labs(x = "y1 (63.05%)", y = "y3 (9.383%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  # theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "none")

####################
```

```r
# change legend position for final plot

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

get_legend<-function(myggplot){
  tmp <- ggplot_gtable(ggplot_build(myggplot))
  leg <- which(sapply(tmp$grobs, function(x) x$name) == "guide-box")
  legend <- tmp$grobs[[leg]]
  return(legend)
}

legend <- get_legend(K3_SCALED_WARD.D2_DIM1_DIM2)

K3_SCALED_WARD.D2_DIM1_DIM2 <- K3_SCALED_WARD.D2_DIM1_DIM2 +
                                theme(legend.position = "none")

###################
```

```r
# Combine 2 scatterplots, WARD.D2, K=3

# Input Scaled Crime 2018 Features

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(K3_SCALED_WARD.D2_DIM1_DIM2,
             K3_SCALED_WARD.D2_DIM1_DIM3,
             legend,
             ncol=2,
             nrow= 2,
             layout_matrix = rbind(c(1,2), c(3,3)),
             widths = c(2.7, 2.7), heights = c(2.5, 0.2),
             top = text_grob("Ward, k=3, Input Standardized Crime 2018, Plotted on y1, y2, y3",
                             color = "black", face = "bold", size = 14))

##########################################################################
##########################################################################
```

```r
# Agglomerative Clustering

# WARD.D2

# K = 3

# PC1, PC2, PC3  from Scaled Crime 2018 Features

# Scatterplots on PC1, PC2, PC3 of cluster assignments

#####################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

####################
```

```
# 3 group WARD.D2 scatterplot with CITY labels, input PC1, PC2, PC3
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_PC1_PC2_PC3_WARD.D2_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN, aes(x=y1, y=y2,
                                           color=K3_PC1_PC2_PC3_WARD.D2)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  #  ggtitle("Ward, k=3, Input y1, y2, y3") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  theme(plot.title = element_text(hjust = 0.5)) +
  #  theme(legend.position = "none")
  theme(legend.position = "bottom")

####################
```

```
# 3 group WARD.D2 scatterplot with CITY labels, input PC1, PC2, PC3
# Plotted on PC1, PC3

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_PC1_PC2_PC3_WARD.D2_DIM1_DIM3 <- ggplot(CRIME_2018_ASSIGN, aes(x=y1, y=y3,
                                           color=K3_PC1_PC2_PC3_WARD.D2)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  #  ggtitle("Ward, k=3, Input y1, y2, y3") +
  labs(x = "y1 (63.05%)", y = "y3 (9.383%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  theme(plot.title = element_text(hjust = 0.5)) +
  # theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "none")


####################
```

```
# change legend position for final plot

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

get_legend<-function(myggplot){
  tmp <- ggplot_gtable(ggplot_build(myggplot))
  leg <- which(sapply(tmp$grobs, function(x) x$name) == "guide-box")
  legend <- tmp$grobs[[leg]]
  return(legend)
}

legend <- get_legend(K3_PC1_PC2_PC3_WARD.D2_DIM1_DIM3)

K3_PC1_PC2_PC3_WARD.D2_DIM1_DIM3 <- K3_PC1_PC2_PC3_WARD.D2_DIM1_DIM3 +
                                    theme(legend.position = "none")

####################
```

```
# Combine 2 scatterplots, WARD.D2, K=3

# Input PC1,PC2, PC3 from Scaled Crime 2018 Features

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(KM3_PC1_PC2_PC3_DIM1_DIM2,
             KM3_PC1_PC2_PC3_DIM1_DIM3,
             legend,
             ncol=2,
             nrow=2,
             layout_matrix = rbind(c(1,2), c(3,3)),
             widths = c(2.7, 2.7), heights = c(2.5, 0.2),
             top = text_grob("Ward, k=3, Input y1, y2, y3, Plotted on y1, y2, y3",
                             color = "black", face = "bold", size = 14))

####################################################################
####################################################################
####################################################################
```

```
# Agglomerative Clustering

# WARD.D2

# k = 3

# Standardized Crime 2018 Features

# Scattermatrix on original dimensions

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

##############################
```

```
# Scaled Crime 2018 Features

# ggpairs

# k = 3

# WARD.D2

p <- ggpairs(CRIME_2018_ASSIGN, c(2,3,4,5,6,7,8,14),
             mapping = ggplot2::aes_string(color = "K3_SCALED_WARD.D2"))

for(i in 1:p$nrow) {
  for(j in 1:p$ncol){
    p[i,j] <- p[i,j] +
      scale_fill_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
      scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788"))
  }
}

p

####################################################################
####################################################################
```

```r
# Agglomerative Clustering

# WARD.D2

# k = 3

# PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# Scattermatrix on original dimensions

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

##############################
```

```r
# PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# ggpairs

# k = 3

# WARD.D2

p <- ggpairs(CRIME_2018_ASSIGN, c(2,3,4,5,6,7,8,15),
             mapping = ggplot2::aes_string(color = "K3_PC1_PC2_PC3_WARD.D2"))

for(i in 1:p$nrow) {
  for(j in 1:p$ncol){
    p[i,j] <- p[i,j] +
      scale_fill_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
      scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788"))
  }
}

p

######################################################################
######################################################################
######################################################################
######################################################################
```

```
# Agglomerative Clustering

# AVERAGE

# Scaled Crime 2018 Features

# Estimating the optimal number of clusters

# NbClust Method

#####################

# install.packages("factoextra")
# library("factoextra")

# install.packages("NbClust", dependencies = TRUE)
# library("NbClust")

# install.packages("magrittr", dependencies = TRUE)
# library("magrittr")

# install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

####################
```

```
# Scaled Crime 2018 Features

# install.packages("NbClust", dependencies = TRUE)
# library("NbClust")

NBCLUST_SCALE_CRIME_2018_FEAT_AVERAGE <- NbClust(SCALED_CRIME_2018_FEAT,
                                                 distance = "euclidean",
                                                 min.nc = 2,
                                                 max.nc = 6,
                                                 method = "average")

# install.packages("factoextra", dependencies = TRUE)
# library("factoextra")

NBCLUST_SCALE_CRIME_2018_FEAT_AVERAGE_BOX <- fviz_nbclust(NBCLUST_SCALE_CRIME_2018_FEAT_AVERAGE
                                                          barfill = "steelblue",
                                                          barcolor = "steelblue") +
  labs(title= "NbClust, Black-Box Method, Average, Input Standardized US Crime 2018") +
  xlab("# of clusters k") +
  ylab("Freq. Among Indices") +
  theme(plot.title = element_text(hjust = 0.5))

####################################################################
####################################################################
```

```
# Agglomerative Clustering

# AVERAGE

# Estimating the optimal number of clusters

# PC1, PC2, PC3 from Scaled Crime 2018 Features

######################

# install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")

# install.packages("factoextra")
# library("factoextra")

# install.packages("NbClust", dependencies = TRUE)
# library("NbClust")

#####################
```

```
# PC1, PC2, PC3 from Scaled Crime 2018 Features

# install.packages("NbClust", dependencies = TRUE)
# library("NbClust")

NBCLUST_PC1_PC2_PC3_CRIME_2018_AVERAGE <- NbClust(PC1_PC2_PC3_CRIME,
                                                  distance = "euclidean",
                                                  min.nc = 2,
                                                  max.nc = 6,
                                                  method = "average")

# install.packages("factoextra", dependencies = TRUE)
# library("factoextra")

NBCLUST_PC1_PC2_PC3_CRIME_2018_AVERAGE_BOX <- fviz_nbclust(NBCLUST_PC1_PC2_PC3_CRIME_2018_AVERAGE
                                                           barfill = "steelblue",
                                                           barcolor = "steelblue") +
  labs(title= "NbClust, Black-Box Method, Average, Input y1, y2, y3") +
  xlab("# of clusters k") +
  ylab("Freq. Among Indices") +
  theme(plot.title = element_text(hjust = 0.5))

#########################################################
```

```
# grid of plots for OPTIMAL K

# install.packages("magrittr", dependencies = TRUE)
# library("magrittr")
# install.packages("ggpubr", dependencies = TRUE)
# library("ggpubr")

OPTIMAL_AVERAGE_CRIME_2018 <- ggarrange(NBCLUST_SCALE_CRIME_2018_FEAT_AVERAGE_BOX,
                                        NBCLUST_PC1_PC2_PC3_CRIME_2018_AVERAGE_BOX,
                                        ncol = 1,
                                        nrow = 2)

######################################################################
######################################################################
######################################################################
```

```
# Agglomerative Clustering

#  AVERAGE

# Compute algorithm

# Scaled Crime 2018 Features

# PC1, PC2, PC3 from Scaled Crime 2018 Features

######################

# Compute "average" using Scaled Crime 2018 Features

set.seed(153)
SCALED_AVERAGE <- hclust(d = DIST_SCALED_CRIME, method = "average")

# Compute "average" using PC1, PC2, PC3

set.seed(154)
PC1_PC2_PC3_AVERAGE <- hclust(d = DIST_PC1_PC2_PC3_CRIME, method = "average")

####################################################################
####################################################################
####################################################################
```

```
# Agglomerative Clustering

# AVERAGE

# K = 3

# add cluster assignments

# Scaled Crime 2018 Features

# PC1, PC2, PC3 from Scaled Crime 2018 Features

######################

#install.packages("plyr", dependencies = TRUE)
#library("plyr")

#####################
```

```
# Cut tree into 3 clusters "AVERAGE" from Scaled Crime 2018 Features

K3_SCALED_AVERAGE <- as.character(cutree(SCALED_AVERAGE, k = 3))

table(K3_SCALED_AVERAGE)

# Cut tree into 3 clusters "AVERAGE" from PC1, PC2, PC3

K3_PC1_PC2_PC3_AVERAGE <- as.character(cutree(PC1_PC2_PC3_AVERAGE, k = 3))

table(K3_PC1_PC2_PC3_AVERAGE)

#install.packages("plyr", dependencies = TRUE)
#library("plyr")

K3_SCALED_AVERAGE <- revalue(K3_SCALED_AVERAGE, c("1"="3", "2"="2", "3"="1"))
K3_PC1_PC2_PC3_AVERAGE <- revalue(K3_PC1_PC2_PC3_AVERAGE, c("1"="3", "2"="2", "3"="1"))

# create dataframe to combine all the data and results

CRIME_2018_ASSIGN <- as.data.frame(cbind(
  CRIME_2018_ASSIGN,
  K3_SCALED_AVERAGE,
  K3_PC1_PC2_PC3_AVERAGE)
)

CRIME_2018_ASSIGN

# set rownames CITY_CRIME_2018

row.names(CRIME_2018_ASSIGN) <- CRIME_2018_ASSIGN$CITY
```

```
# summary

summary(CRIME_2018_ASSIGN)

####################################################################
####################################################################
####################################################################

# AVERAGE

# k = 3

# Scaled Crime 2018 Features

# PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# Cluster City Names, Cluster Mean Vectors, Cluster sd Vectors,
# check for differences in cluster assignments

#####################
```

```
# cluster city names

# AVERAGE 3 cluster solution on Scaled Features 2018

# cluster 1

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_SCALED_AVERAGE == "1",1])

# cluster 2

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_SCALED_AVERAGE == "2",1])

# cluster 3

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_SCALED_AVERAGE == "3",1])

###########################
```

```
# cluster city names

# AVERAGE 3 cluster solution on PC1, PC2, PC3 derived from Scaled Crime 2018 Features

# cluster 1

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_PC1_PC2_PC3_AVERAGE == "1",1])

# cluster 2

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_PC1_PC2_PC3_AVERAGE == "2",1])

# cluster 3

sort(CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$K3_PC1_PC2_PC3_AVERAGE == "3",1])

####################

# 1-way table of cluster assignments

table(K3_SCALED_AVERAGE)

table(K3_PC1_PC2_PC3_AVERAGE)

#####################

# 2-way table of cluster assignments (confusion matrix)

table(K3_SCALED_AVERAGE, K3_PC1_PC2_PC3_AVERAGE)

#####################
```

```
# cluster means, AVERAGE, k=3, standardized Crime 2018 Features

aggregate(CRIME_2018_ASSIGN[, 2:8], list(CRIME_2018_ASSIGN$K3_SCALED_AVERAGE), mean)

###########################

# cluster means, AVERAGE, k=3, PC1, PC2, PC3 derived from Scaled Crime 2018 Features

aggregate(CRIME_2018_ASSIGN[, 2:8], list(CRIME_2018_ASSIGN$K3_PC1_PC2_PC3_AVERAGE), mean)

##########################

#install.packages("psych", dependencies = TRUE)
# library("psych")

# original sample means

describe(CRIME_2018_FEAT)

####################################################################
####################################################################
####################################################################
```

```
# Agglomerative Clustering

# AVERAGE

# K = 3

# Dendrograms

# Scaled Crime 2018 Features

#######################

# install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")

# install.packages("factoextra")
# library("factoextra")

#######################
```

```
# AVERAGE, Input Scaled Crime 2018, Rectangle Dendrogram, K=3

DEND_RECTANGLE_K3_SCALED_AVERAGE <- fviz_dend(SCALED_AVERAGE, k = 3, # Cut in three groups
                                              cex = 0.5, # label size
                                              palette = c("#009788", "#FE3A07", "#0080FF"),
                                              color_labels_by_k = TRUE, # color labels by groups
                                              rect = TRUE, # Add rectangle around groups
                                              # main = NULL,
                                              main = "Average, k=3, Input Standardized
                                              Crime 2018, Rectangular Dendrogram",
                                              rect_border = c("#009788", "#FE3A07", "#0080FF"),
                                              rect_fill = TRUE) +
                                              theme(plot.title = element_text(hjust = 0.5))

####################################################################
####################################################################
####################################################################
```

```
# Agglomerative Clustering

# AVERAGE

# K = 3

# Dendrograms

# PC1, PC2, PC3 from Scaled Crime 2018 Features

######################

# install.packages("ggplot2", dependencies = TRUE)
# library("ggplot2")

# install.packages("factoextra")
# library("factoextra")

######################
```

```
# AVERAGE, Input PC1, PC2, PC3, Rectangle Dendrogram, K=3

DEND_RECTANGLE_K3_PC1_PC2_PC3_AVERAGE <- fviz_dend(PC1_PC2_PC3_AVERAGE, k = 3, # Cut in three groups
                                                   cex = 0.5, # label size
                                                   palette = c("#009788", "#FE3A07", "#0080FF"),
                                                   color_labels_by_k = TRUE, # color labels by groups
                                                   rect = TRUE, # Add rectangle around groups
                                                   main = "Average, k=3, Input y1, y2, y3,
                                                   Rectangular Dendrogram",
                                                   rect_border = c("#009788", "#FE3A07", "#0080FF"),
                                                   rect_fill = TRUE) +
                                                  theme(plot.title = element_text(hjust = 0.5))

######################################################################
######################################################################
######################################################################
```

```
# AVERAGE

# K = 3

# Scaled Crime 2018 Features

# Scatterplots on PC1, PC2, PC3 of cluster assignments and ggpairs matrix

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

#####################

# 3 group AVERAGE scatterplot with CITY labels input scaled crime 2018 data.
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

####################
```

```
# 3 group AVERAGE scatterplot with CITY labels input scaled crime 2018 data.
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_SCALED_AVERAGE_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN,
                                      aes(x=y1, y=y2, color=K3_SCALED_AVERAGE)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  #  ggtitle("Average, k=3, Input Scaled Crime 2018") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c( "#FE3A07", "#0080FF", "#009788")) +
  #  theme(plot.title = element_text(hjust = 0.5)) +
  #  theme(legend.position = "none")
  theme(legend.position = "bottom")

####################
```

```
# 3 group AVERAGE scatterplot with CITY labels input scaled crime 2018 data.
# Plotted on PC1, PC3

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_SCALED_AVERAGE_DIM1_DIM3 <- ggplot(CRIME_2018_ASSIGN,
                                      aes(x=y1, y=y3, color=K3_SCALED_AVERAGE)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  # ggtitle("Average, k=3, Input Scaled Crime 2018") +
  labs(x = "y1 (63.05%)", y = "y3 (9.383%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  # theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "none")

####################
```

```
# change legend position for final plot

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

get_legend<-function(myggplot){
  tmp <- ggplot_gtable(ggplot_build(myggplot))
  leg <- which(sapply(tmp$grobs, function(x) x$name) == "guide-box")
  legend <- tmp$grobs[[leg]]
  return(legend)
}

legend <- get_legend(K3_SCALED_AVERAGE_DIM1_DIM2)

K3_SCALED_AVERAGE_DIM1_DIM2 <- K3_SCALED_AVERAGE_DIM1_DIM2 +
                               theme(legend.position = "none")

####################
```

```
# Combine 2 scatterplots, AVERAGE, K=3

# Input Scaled Crime 2018 Features

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(K3_SCALED_AVERAGE_DIM1_DIM2,
             K3_SCALED_AVERAGE_DIM1_DIM3,
             legend,
             ncol=2,
             nrow= 2,
             layout_matrix = rbind(c(1,2), c(3,3)),
             widths = c(2.7, 2.7), heights = c(2.5, 0.2),
             top = text_grob("Average, k=3, Input Standardized Crime 2018,
                             Plotted on y1, y2, y3",
                             color = "black", face = "bold", size = 14))

##############################################################################
##############################################################################
```

```
# Agglomerative Clustering

# AVERAGE

# K = 3

# PC1, PC2, PC3  from Scaled Crime 2018 Features

# Scatterplots on PC1, PC2, PC3 of cluster assignments

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

#####################
```

```
# 3 group AVERAGE scatterplot with CITY labels, input PC1, PC2, PC3
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_PC1_PC2_PC3_AVERAGE_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN,
                                           aes(x=y1, y=y2, color=K3_PC1_PC2_PC3_AVERAGE)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  #  ggtitle(Average, k=3, Input y1, y2, y3") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  theme(plot.title = element_text(hjust = 0.5)) +
  #  theme(legend.position = "none")
  theme(legend.position = "bottom")

###################
```

```
# 3 group AVERAGE scatterplot with CITY labels, input PC1, PC2, PC3
# Plotted on PC1, PC3

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_PC1_PC2_PC3_AVERAGE_DIM1_DIM3 <- ggplot(CRIME_2018_ASSIGN,
                                           aes(x=y1, y=y3, color=K3_PC1_PC2_PC3_AVERAGE))
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  #  ggtitle("Average, k=3, Input y1, y2, y3") +
  labs(x = "y1 (63.05%)", y = "y3 (9.383%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "none")

####################
```

```
# change legend position for final plot

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

get_legend<-function(myggplot){
  tmp <- ggplot_gtable(ggplot_build(myggplot))
  leg <- which(sapply(tmp$grobs, function(x) x$name) == "guide-box")
  legend <- tmp$grobs[[leg]]
  return(legend)
}

legend <- get_legend(K3_PC1_PC2_PC3_AVERAGE_DIM1_DIM2)

K3_PC1_PC2_PC3_AVERAGE_DIM1_DIM2 <- K3_PC1_PC2_PC3_AVERAGE_DIM1_DIM2 +
                                    theme(legend.position = "none")

####################
```

```
# Combine 2 scatterplots, AVERAGE, K=3

# Input PC1,PC2, PC3 from Scaled Crime 2018 Features

#install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

grid.arrange(K3_PC1_PC2_PC3_AVERAGE_DIM1_DIM2,
             K3_PC1_PC2_PC3_AVERAGE_DIM1_DIM3,
             legend,
             ncol=2,
             nrow=2,
             layout_matrix = rbind(c(1,2), c(3,3)),
             widths = c(2.7, 2.7), heights = c(2.5, 0.2),
             top = text_grob("Average, k=3, Input y1, y2, y3,
                             Plotted on y1, y2, y3",
                             color = "black", face = "bold", size = 14))

####################################################################
####################################################################
####################################################################
```

```
# Agglomerative Clustering

# AVERAGE

# k = 3

# Standardized Crime 2018 Features

# Scattermatrix on original dimensions

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

##############################
```

```
# Scaled Crime 2018 Features

# ggpairs

# k = 3

# AVERAGE

p <- ggpairs(CRIME_2018_ASSIGN, c(2,3,4,5,6,7,8,16),
             mapping = ggplot2::aes_string(color = "K3_SCALED_AVERAGE"))

for(i in 1:p$nrow) {
  for(j in 1:p$ncol){
    p[i,j] <- p[i,j] +
      scale_fill_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
      scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788"))
  }
}

p

####################################################################
####################################################################
# Agglomerative Clustering

# AVERAGE

# k = 3

# PC1,PC2, PC3 from Scaled Crime 2018 Features

# Scattermatrix on original dimensions

######################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

##############################
```

```
# PC1,PC2, PC3 from Scaled Crime 2018 Features

# ggpairs

# k = 3

# AVERAGE

p <- ggpairs(CRIME_2018_ASSIGN, c(2,3,4,5,6,7,8,17),
             mapping = ggplot2::aes_string(color = "K3_PC1_PC2_PC3_AVERAGE"))

for(i in 1:p$nrow) {
  for(j in 1:p$ncol){
    p[i,j] <- p[i,j] +
      scale_fill_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
      scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788"))
  }
}

p

####################################################################
####################################################################
####################################################################
####################################################################
```

```
# AVERAGE, WARD.D2

# Scaled Crime 2018 Features

# PC1,PC2, PC3 from Scaled Crime 2018 Features

# Compare Common Dendrograms Branches using Tanglegrams

####################

# install.packages("dendextend", dependencies = TRUE)
library("dendextend")

####################
```

```
# create dendrograms that work with "dendextend"

# ward.d2

# scaled crime 2018

DEND_SCALE_WARD.D2 <- as.dendrogram(SCALED_WARD.D2)

# pc1, pc2, pc3

DEND_PC1_PC2_PC3_WARD.D2 <- as.dendrogram(PC1_PC2_PC3_WARD.D2)

# average

# scaled crime 2018

DEND_SCALE_AVERAGE <- as.dendrogram(SCALED_AVERAGE)

# pc1, pc2, pc3

DEND_PC1_PC2_PC3_AVERAGE <- as.dendrogram(PC1_PC2_PC3_AVERAGE)

########################
```

```
# tanglegram

TANGLE_WARD_SCALE_PC1_PC2_PC3 <- tanglegram(dend1 = DEND_SCALE_WARD.D2,
                                            dend2 = DEND_PC1_PC2_PC3_WARD.D2,
                                            common_subtrees_color_branches = TRUE,
                                            main_left = "Ward, Input S. Crime 2018",
                                            main_right = "Ward, Input y1, y2, y3"
)

TANGLE_AVERAGE_SCALE_PC1_PC2_PC3 <- tanglegram(dend1 = DEND_SCALE_AVERAGE,
                                               dend2 = DEND_PC1_PC2_PC3_AVERAGE,
                                               common_subtrees_color_branches = TRUE,
                                               main_left = "Average, Input S. Crime 2018",
                                               main_right = "Average, Input y1, y2, y3"
)

TANGLE_WARD_AVERAGE_SCALE <- tanglegram(dend1 = DEND_SCALE_WARD.D2,
                                        dend2 = DEND_SCALE_AVERAGE,
                                        common_subtrees_color_branches = TRUE,
                                        main_left = "Ward, Input S. Crime 2018",
                                        main_right = "Average, Input S. Crime 2018"
)

TANGLE_WARD_AVERAGE_PC1_PC2_PC3 <- tanglegram(dend1 = DEND_PC1_PC2_PC3_WARD.D2,
                                              dend2 = DEND_PC1_PC2_PC3_AVERAGE,
                                              common_subtrees_color_branches = TRUE,
                                              main_left = "Ward, Input y1, y2, y3",
                                              main_right = "Average, Input y1, y2, y3",
)
```

```
######################################################################
######################################################################
######################################################################
######################################################################

# K-means, WARD.D2, AVERAGE

# K=3

# Scaled Crime 2018 Features

# PC1,PC2, PC3 from Scaled Crime 2018 Features

# compare scatterplots cluster assignments

#####################

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

#install.packages("GGally", dependencies = TRUE)
#library("GGally")

# install.packages("gridExtra", dependencies = TRUE)
# library("gridExtra")

#####################
```

```r
# 3 group k-mean scatterplot with CITY labels input standardized crime 2018 data.
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

KM3_SCALE_CRIME_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN,
                                    aes(x=y1, y=y2, color=KM3_SCALED_ASSIGN)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  ggtitle("k-Means, k=3, Input Standardized Crime 2018") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF","#009788")) +
  theme(plot.title = element_text(hjust = 0.5))
# theme(legend.position = "none")
# theme(legend.position = "bottom")

###############################
```

```r
# 3 group k-mean scatterplot with CITY labels, input PC1, PC2, PC3
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

KM3_PC1_PC2_PC3_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN,
                                    aes(x=y1, y=y2, color=KM3_PC1_PC2_PC3_ASSIGN)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  ggtitle("k-Means, k=3, Input y1, y2, y3") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  theme(plot.title = element_text(hjust = 0.5))
# theme(legend.position = "none")
# theme(legend.position = "bottom")


###############################
```

```r
###############################

# 3 group WARD.D2 scatterplot with CITY labels input scaled crime 2018 data.
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_SCALED_WARD.D2_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN,
                                      aes(x=y1, y=y2, color=K3_SCALED_WARD.D2)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  ggtitle("Wards, k=3, Input Standardized Crime 2018") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c( "#FE3A07", "#0080FF", "#009788")) +
  theme(plot.title = element_text(hjust = 0.5))
# theme(legend.position = "none")
# theme(legend.position = "bottom")

###############################
```

```r
# 3 group WARD.D2 scatterplot with CITY labels, input PC1, PC2, PC3
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_PC1_PC2_PC3_WARD.D2_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN,
              aes(x=y1, y=y2, color=K3_PC1_PC2_PC3_WARD.D2)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  ggtitle("Ward, k=3, Input y1, y2, y3") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  theme(plot.title = element_text(hjust = 0.5))
# theme(legend.position = "none")
# theme(legend.position = "bottom")

############################
```

```r
# 3 group AVERAGE scatterplot with CITY labels input scaled crime 2018 data.
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_SCALED_AVERAGE_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN,
             aes(x=y1, y=y2, color=K3_SCALED_AVERAGE)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  ggtitle("Average, k=3, Input Scaled Crime 2018") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c( "#FE3A07", "#0080FF", "#009788")) +
  theme(plot.title = element_text(hjust = 0.5))
# theme(legend.position = "none")
# theme(legend.position = "bottom")

##############################
```

```r
# 3 group AVERAGE scatterplot with CITY labels, input PC1, PC2, PC3
# Plotted on PC1, PC2

#install.packages("ggplot2", dependencies = TRUE)
#library("ggplot2")

K3_PC1_PC2_PC3_AVERAGE_DIM1_DIM2 <- ggplot(CRIME_2018_ASSIGN,
                aes(x=y1, y=y2, color=K3_PC1_PC2_PC3_AVERAGE)) +
  geom_text(label=rownames(CRIME_2018_ASSIGN)) +
  ggtitle("Average, k=3, Input y1, y2, y3") +
  labs(x = "y1 (63.05%)", y = "y2 (10.99%)", col = "cluster") +
  scale_color_manual(values=c("#FE3A07", "#0080FF", "#009788")) +
  theme(plot.title = element_text(hjust = 0.5))
# theme(legend.position = "none")
# theme(legend.position = "bottom")

#################################
```

```
# combined plots

# Scaled Crime 2018 Features

# PC1,PC2, PC3 from Scaled Crime 2018 Features

grid.arrange(
  KM3_SCALE_CRIME_DIM1_DIM2,
  KM3_PC1_PC2_PC3_DIM1_DIM2,
  K3_SCALED_WARD.D2_DIM1_DIM2,
  K3_PC1_PC2_PC3_WARD.D2_DIM1_DIM2,
  K3_SCALED_AVERAGE_DIM1_DIM2,
  K3_PC1_PC2_PC3_AVERAGE_DIM1_DIM2,
  nrow = 3) #,
# top = text_grob("K=3, K-Means Row 1, WARD Row 2, AVERAGE Row 3",
#                 color = "black", face = "bold", size = 14))

#####################################
```

```
# K=3, KMEANs, Cluster assignments

summary(CRIME_2018_ASSIGN[,c(12,13)])

# K=3, WARD, Cluster assignments

summary(CRIME_2018_ASSIGN[,c(14,15)])

# K=3, AVERAGE, Cluster assignments

summary(CRIME_2018_ASSIGN[,c(16,17)])

####################################

CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$KM3_SCALED_ASSIGN == "1",]

###################################################################
###################################################################
###################################################################
###################################################################
```

```
# conclusion and future study

High_Crime <- c("Albuquerque",
                "Anchorage",
                "Baltimore",
                "Chicago",
                "Detroit",
                "Houston",
                "Lake_Charles",
                "Little_Rock",
                "Memphis",
                "Myrtle_Beach",
                "Nashville",
                "New_Orleans",
                "St_Louis")
High_Crime

# original crime variables

CRIME_2018_ASSIGN[CRIME_2018_ASSIGN$CITY == High_Crime,2:8]

# y1, y2, y3

CRIME_2018_ASSIGN[High_Crime,9:11]
```