Multivariate Techniques

Multivariate Techniques:

An Example Based Approach

By

Muhammad Qaiser Shahbaz,
Saman Hanif Shahbaz
and Muhammad Hanif

Cambridge
Scholars
Publishing

Multivariate Techniques: An Example Based Approach

By Muhammad Qaiser Shahbaz, Saman Hanif Shahbaz
and Muhammad Hanif

This book first published 2019

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-4011-1
ISBN (13): 978-1-5275-4011-8

To My Sisters
Muhammad Qaiser Shahbaz

To My Sons: Khizar, Yousuf and Mussa
Saman Hanif Shahbaz

To My Parents
Muhammad Hanif

CONTENTS

PREFACE

Multivariate analysis has tremendous applications in many areas of life and deals with several random variables simultaneously. The data which arise in practical life are multivariate in nature and hence require specialized techniques for decision making. Multivariate analyses are studied in two contexts, namely the multivariate distributions and multivariate techniques. Both of these contexts have their own requirements. The multivariate distributions are mainly studied theoretically whereas the multivariate techniques are studied using real data.

In this we have given a detailed insight of multivariate distributions and multivariate techniques. The opening chapter of the book provides an introduction to multivariate distribution theory and popular multivariate measures. This chapter also provides a brief review of matrix algebra. One of the most popular multivariate distributions, the multivariate normal distribution, has been discussed in detail in chapter 2. Some other popular multivariate distributions have also been discussed in chapter 3. These include the Wishart distribution, multivariate Beta distribution and Wilk's distribution.

Chapter 4 and Chapter 5 contain two popular multivariate inferences, namely inferences about population mean vectors and population covariance matrices. The multivariate counterpart of students' t statistics, the Hotelling's T^2 statistic has also been discussed in Chapter 4. The multivariate extension of popular Analysis of Variance (*ANOVA*), the multivariate *ANOVA*, has been given in Chapter 6. This chapter contains both one way and two way multivariate *ANOVA* and also provides methods to conduct inference about several mean vectors.

Chapter 7 of the book is dedicated to another popular multivariate technique, the multivariate regression analysis. This chapter provides methods of estimation and hypotheses testing in case of multivariate regression analysis. The chapter also covers another popular multivariate technique, the Seemingly Unrelated Regression (*SUR*) models and contains methods of estimation and hypothesis testing for such a model.

Chapter 8 and Chapter 9 are dedicated to multivariate techniques which do not have univariate counterpart. Chapter 8 is reserved for a detailed discussion of one of these techniques, the Canonical Correlation Analysis (*CCA*). Various versions of the technique have been discussed including simple, partial, part and bi–partial canonical correlations. Methods of computing these correlations and variate are discussed in Chapter 8 alongside the inferences in these variants of canonical correlation.

Chapter 9 of the book is dedicated to two popular dimension reduction techniques, the Principal Component Analysis (*PCA*) and Factor Analysis (*FA*). Methods of computing principal components and factors are discussed alongside the procedures to conduct inferences in these techniques.

The book provides sufficient numerical examples to understand various concepts alongside *R* codes for these examples.

Finally, we would like to thank our colleagues and students for critical comments throughout compilation of this book. Author one would like to thank his elder brother Prof. Dr. Tariq Bhatti for his continuous support and guidance in his academic and personal life. Authors one and two would also like to thank Department of Statistics, King Abdulaziz University for providing excellent facilities to compile this book, whereas the third author is thankful to National College of Business Administration & Economics, Lahore for providing a good atmosphere to work on this book.

Muhammad Qaiser Shahbaz, Saman Hanif Shahbaz
and Muhammad Hanif

CHAPTER ONE

INTRODUCTION

1.1 Introduction

The data always arise in scientific and social studies as an input to the study. The studies are conducted to achieve certain objectives. In some cases the objectives are such that the data is required on a single variable only, for example study related to life length of an electronic component. When the data is available on a single variable, the univariate analyses are very helpful. For example we can use the popular t-test to test the hypothesis that the average life length of an electric component is equal to 2 years. In certain situations, data on several variables is available and is called the *multivariate data*. If these variables are mutually independent then each variable can be studied separately but this is a rare case. In most of the situations, the variables are dependent and hence can not be studied individually by using univariate analysis of each variable separately. In such situations we need techniques which help us in studying data on several variables and are called *multivariate techniques*. Further, the analysis of multivariate data is called *multivariate analysis*. In this book, we have discussed some popular multivariate techniques to analyze a multivariate data depending upon various objectives. We have discussed these techniques in two ways, namely the theory of the techniques and their applications on the real data.

In some cases, the multivariate analyses are a straight extension of the univariate analyses and in some cases these analyses are stand alone analyses and do not have any univariate counterpart. We have discussed both type of techniques in this book.

The multivariate analysis require certain specialized terms and notations which we will discuss in this chapter. These terms and notations are useful in understanding and applying the multivariate analysis. We start with a basic cornerstone of the multivariate analysis, known as the data matrix in the following.

1.2 The Data Matrix

The data is a key input to statistical analysis. In univariate analysis, the data is collected from a set of respondents on a single variable and is usually presented in the form of a column vector. These vectors are then analyzed with respect to ceratian underlying objectives. In multivariate analysis the data is collected from a set of respondents on a set of variables and is usually collected in the form of a matrix. The matrix which contains the multivariate data is called the *data matrix* and has following description

$$\mathbf{X}_{(p\times n)} = \begin{bmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1j} & \cdots & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2j} & \cdots & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & \cdots & x_{ij} & \cdots & \cdots & x_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & \cdots & \cdots & x_{pj} & \cdots & \cdots & x_{pn} \end{bmatrix};$$

$$= \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \cdots & \mathbf{x}_j & \cdots & \cdots & \mathbf{x}_n \end{bmatrix}.$$

In above representation of the data matrix each column vector, $\mathbf{x}_j$, is collection of information of all variables for *j*th respondent and hence is *j*th observation in a multivariate data. The data matrix is a key sample information in *Multivariate Analysis* as this is needed to compute almost all of the multivariate measures. We will see this in the coming chapters.

We know that in univariate analysis the study can be done by using the data and by using the distribution of underlying data. The study using data is usually considered a sample based study but often it happended that the study is conducted by using underlying probability distribution from where the data has been drawn. Such studies are conducted on the basis of some random variables. In multivariate analysis the concept is extended to the collection of several random variables. This collection of random variables is usually represented in the form of a vector is and called the random vector which is discussed below.

1.2.1 The Random Vector

In univariate analysis, a random variable plays very important role as it provides information about probability distribution of some phenomenon. A random variable X always has some distribution function $F(x)$, which provide all the information about that random variable and is known as *univariate distribution*. In univariate analysis, the sample is drawn from some univariate distribution and is studied based upon the underlying objectives. In multivariate analysis concept of a single random variable is extended to the case of several random variables having some joint distribution. In multivariate analysis we have a collection of p random variables, X_1, X_2, … X_p, which can be collected in the form of a column vector $\mathbf{x}$ as

$$\mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_p \end{bmatrix}.$$

The above vector is called a random vector and its ith component X_i is a random variable. Every random vector has certain joint distribution function given as $F(x_1, x_2, \ldots, x_p)$ or $F(\mathbf{x})$. The distribution of a random vector is called a *multivariate distribution*. The random vector is key to all inferences in *multivariate theory*. The random vector is easily extended to the case of a random matrix and is a $(p \times p)$ matrix $\mathbf{X}$ such that all of its entries are are random variables. The distribution of a random matrix is called a *matrix distribution.*

The multivariate distributions require certain notations which appear frequently. In the following we will discuss some notations which appear in multivariate analysis.

1.3 Notations of Bivariate and Multivariate Distributions

Some popular notations which appear in multivariate analysis are discussed in the following.

1.3.1 The Joint Distributions

The univariate distribution is a probability distribution of a single random variable. The concept is easily extended to the case of two and several random variables. We will discuss both in the following.

The joint density function of two continuous random variables X_1 and X_2 is denoted by $f(x_1,x_2)$ so that $f(x_1,x_2)\geq 0$ and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2)\,dx_1dx_2 = 1.$$

The joint distribution function, in case of two random variables, is obtained as

$$F(x_1,x_2)=\int_{-\infty}^{x_2}\int_{-\infty}^{x_1} f(u_1,u_2)\,du_1du_2\,\cdot \tag{1.3.1}$$

The function $F(x_1,x_2)$ is a distribution function if it satisfies following conditions

$$F(+\infty,+\infty)=1, F(-\infty,x_2)=0, F(x_1,-\infty)=0$$

and for every $a_1<a_2$ and $b_1<b_2$ the following inequality holds

$$F(a_2,b_2)+F(a_1,b_1)-F(a_1,b_2)-F(a_2,b_1)\geq 0\cdot$$

The joint distribution function is a udeful function which provides all the information about two random variables. The joint density function of two random variables is easily obtained from the joint distribution function by differentiation, that is the joint density function of two random variables X_1 and X_2 is obtained from the joint distribution function as

$$\frac{\partial^2}{\partial x_1\partial x_2}F(x_1,x_2)=f(x_1,x_2)\cdot \tag{1.3.2}$$

The joint density function is a useful function and provides basis to compute joint probabilities for two random variables. Specifically, the probability that the random variables X_1 and X_2 belong to some region E is computed as

$$P\{(X_1,X_2)\in E\}=\iint_E f(x_1,x_2)\,dx_1dx_2\ . \tag{1.3.3}$$

The joint distribution also provides basis for computation of joint moments of two random variables which as

$$\mu'_{r_1,r_2}=E\left(X_1^{r_1}X_2^{r_2}\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1^{r_1}x_2^{r_2}f(x_1,x_2)\,dx_1dx_2\,\cdot \tag{1.3.4}$$

The concept of joint distribution is easily extended to the case of several random variables and the density for several random variables $X_1, X_2, \ldots, X_p$ is denoted by $f\left(x_1, x_2, \ldots, x_p\right)$ such that $f\left(x_1, x_2, \ldots, x_p\right) \geq 0$ and

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f\left(x_1, \ldots, x_p\right) dx_1 \ldots dx_p = 1 \cdot$$

The joint density function of several random variables provide basis for the joint distribution function of p random variables which is given as

$$F\left(x_1, \ldots, x_p\right) = \int_{-\infty}^{x_p} \ldots \int_{-\infty}^{x_1} f\left(u_1, \ldots, u_p\right) du_1 \ldots du_p . \tag{1.3.5}$$

The joint distribution function of serveral random variables satisfies $F\left(+\infty, \ldots, +\infty\right) = 1$ and $F\left(x_1, \ldots, x_{i-1}, -\infty, x_{i+1}, \ldots, x_p\right) = 0$. The joint density function of several random variables is directly obtained from the joint distribution function by differentiation as

$$\frac{\partial^p}{\partial x_1 \ldots \partial x_p} F\left(x_1, \ldots, x_p\right) = f\left(x_1, \ldots, x_p\right) \cdot \tag{1.3.6}$$

The joint density function of several random variables can be used to compute the joint probability for occuence of several random variables. Specifically, the probability that the random variables belong to some measurable space A is computed, by using the joint density function, as

$$P\left\{\left(X_1, \ldots, X_p\right) \in A\right\} = \int \cdots \int_A f\left(x_1, \ldots, x_p\right) dx_1 \ldots dx_p . \tag{1.3.7}$$

The joint density function of several random variables is useful in computing the product moments for several random varibales. Specifically, the product moments are computed by using

$$\mu'_{r_1, \ldots, r_p} = E\left(X_1^{r_1} \ldots X_p^{r_p}\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{r_1} \ldots x_p^{r_p} f\left(x_1, \ldots, x_2\right) dx_1 \ldots dx_2 . \tag{1.3.8}$$

The product moment, given in (1.3.8), is useful in computing moments for single random variables individually.

1.3.2 The Marginal Distributions

The joint distribution of two random variables provides information about joint behavior of two random variables. This joint distribution also provides basis to study each of the random variable individually by obtaining distributions of both random variables seperately. This is illustrated in the following.

Suppose that two random variables X_1 and X_2 has joint density function $f\left(x_1,x_2\right)$. The marginal distributions of X_1 and X_2 are given as

$$f_1\left(x_1\right)=\int_{-\infty}^{\infty}f\left(x_1,x_2\right)dx_2 \text{ and } f_2\left(x_2\right)=\int_{-\infty}^{\infty}f\left(x_1,x_2\right)dx_1 \cdot \tag{1.3.9}$$

The raw marginal moments for random variables X_1 and X_2 are given as

$$E\left(X_1^{r_1}\right)=\int_{-\infty}^{\infty}x_1^{r_1}f_1\left(x_1\right)dx_1 \text{ and } E\left(X_2^{r_2}\right)=\int_{-\infty}^{\infty}x_2^{r_2}f\left(x_2\right)dx_2 \cdot \tag{1.3.10}$$

The central moments for both of the random variables are defined as

$$E\left\{\left(X_1-\mu_1\right)^{r_1}\right\}=\int_{-\infty}^{\infty}\left(x_1-\mu_1\right)^{r_1}f_1\left(x_1\right)dx_1; \tag{1.3.11}$$

and

$$E\left\{\left(X_2-\mu_2\right)^{r_2}\right\}=\int_{-\infty}^{\infty}\left(x_2-\mu_2\right)^{r_2}f\left(x_2\right)dx_2 \cdot \tag{1.3.12}$$

The quantities $E\left\{\left(X_1-\mu_1\right)^2\right\}=\sigma_{11}$ and $E\left\{\left(X_2-\mu_2\right)^2\right\}=\sigma_{22}$ are called the variances of random variables X_1 and X_2. The covariance between X_1 and X_2 is defined as

$$E\left\{\left(X_1-\mu_1\right)\left(X_2-\mu_2\right)\right\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x_1-\mu_1\right)\left(x_2-\mu_2\right)f\left(x_1,x_2\right)dx_1dx_2 \cdot \tag{1.3.13}$$

The joint distribution of several random variables is useful in obtaining the marginal distributions of each of the random variable. In case of the joint distribution of several random variables, the marginal distribution for *i*th random variable is defined as:

$$f_i\left(x_i\right)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f\left(x_1,\ldots,x_p\right)dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_p; \tag{1.3.14}$$

and the joint marginal distribution of a sub-set $X_1, X_2, \ldots, X_q$ is obtained as

$$f\left(x_1,\ldots,x_q\right)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f\left(x_1,\ldots,x_p\right)dx_{q+1}\ldots dx_p \cdot \tag{1.3.15}$$

The raw and central moments for ith random variable is obtained by using

$$E\left(X_i^{r_i}\right)=\int_{-\infty}^{\infty}x_i^{r_i}f_i\left(x_i\right)dx_i, \tag{1.3.16a}$$

$$E\left\{\left(X_i-\mu_i\right)^{r_i}\right\}=\int_{-\infty}^{\infty}\left(x_i-\mu_i\right)^{r_i}f_i\left(x_i\right)dx_i, \tag{1.3.16b}$$

where $\mu_i = E(X_i)$ is mean of ith variable and $\sigma_{ii} = E\left\{(X_i - \mu_i)^2\right\}$ is the variance of X_i. The covariance between X_i and X_h is given as

$$E\{(X_i - \mu_i)(X_h - \mu_h)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i - \mu_i)(x_h - \mu_h) f(x_i, x_h)\,dx_i dx_h \cdot \tag{1.3.16c}$$

The correlation coefficient between X_i and X_h is defined as $\rho_{ih} = \sigma_{ih} / (\sigma_{ii}\sigma_{hh})^{1/2}$. The collection of means, variances and covariances plays very important role in multivariate analysis.

1.3.3 The Conditional Distributions

The joint distribution of two random variables can be used to obtain the distribution of one of the random variable under certain conditions on the other random variable. Such a distribution is called a conditional distribution and is defined below.

Given the joint distribution, $f(x_1, x_2)$, of two random variables X_1 and X_2, the conditional distributions are defined as

$$f(x_1 | X_2 = x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} \text{ and } f(x_2 | X_1 = x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}. \tag{1.3.17}$$

For simplicity we write $f(x_1 | X_2 = x_2) = f(x_1 | x_2)$ and $f(x_2 | X_1 = x_1) = f(x_2 | x_1)$. The conditional probability of say $X_1 \in A$ given $X_2 \in B$ is computed by using

$$P(X_1 \in A | X_2 \in B) = \frac{P(X_1 \in A \cap X_2 \in B)}{P(X_2 \in B)} = \frac{\int_A \int_B f(x_1, x_2)\,dx_2 dx_1}{\int_B f_2(x_2)\,dx_2}.$$

The conditional distributions can be used to compute the conditional moments of the random variables. Specifically the raw conditional moments are defined as

$$E\left(X_1^{r_1} | x_2\right) = \int_{-\infty}^{\infty} x_1^{r_1} f(x_1 | x_2)\,dx_1 \text{ and } E\left(X_2^{r_2} | x_1\right) = \int_{-\infty}^{\infty} x_2^{r_2} f(x_2 | x_1)\,dx_2 \cdot \tag{1.3.18}$$

Specifically $E(X_1 | x_2) = \mu_{1|2}$ and $E(X_2 | x_1) = \mu_{2|1}$ are called the conditional means of X_1 given $X_2 = x_2$ and X_2 given $X_1 = x_1$ respectively. The conditional means are sometime called the *regression functions*. The conditional central moments are defined as

$$E\left\{\left(X_1-\mu_{1|2}\right)^{r_1}|x_2\right\}=\int_{-\infty}^{\infty}\left(x_1-\mu_{1|2}\right)^{r_1}f\left(x_1\,|\,x_2\right)dx_1; \quad (1.3.19)$$

and

$$E\left\{\left(X_2-\mu_{2|1}\right)^{r_2}|x_1\right\}=\int_{-\infty}^{\infty}\left(x_2-\mu_{2|1}\right)^{r_2}f\left(x_2\,|\,x_1\right)dx_2\cdot \quad (1.3.20)$$

The quantities

$$E\left\{\left(X_1-\mu_{1|2}\right)^{2}|x_2\right\}=\sigma_{11.2} \text{ and } E\left\{\left(X_2-\mu_{2|1}\right)|x_1\right\}=\sigma_{22.1}$$

are called the conditional or partial variances of X_1 given $X_2 = x_2$ and X_2 given $X_1 = x_1$ respectively.

When we have the joint distribution of several random variables as $f\left(x_1,...,x_p\right)$ then several conditional distributions can be defined. Specifically the conditional distribution of one set of variables say X_1, X_2, …, X_q given the remaining set of variables is obtained as

$$f\left(x_1,...,x_q\left|x_{q+1},...,x_p\right.\right)=\frac{f\left(x_1,...,x_p\right)}{f\left(x_{q+1},...,x_p\right)}\cdot \quad (1.3.21)$$

The conditional means, conditional or partial variances and conditional or partial covariances can be computed from (1.3.19). The partial variances and covariances provide the basis for computation of the partial correlation coefficients.

Example 1.1: The joint distribution of two random variables X_1 and X_2 is

$$f\left(x_1,x_2\right)=\frac{x_1x_2}{2};0<x_1<x_2<2$$
$$=0 \ \textit{otherwise}.$$

Obtain the expression for $(r,s)th$ joint moment. Obtain two marginal distributions and conditional distribution of X_1 given X_2 alongside the conditional mean and variance.

Solution: The joint $(r,s)th$ moment is given as

$$\mu_{r,s}^{/}=E\left(X_1^rX_2^s\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1^rx_2^sf\left(x_1,x_2\right)dx_1dx_2$$

$$= \frac{1}{2}\int_0^2\int_0^{x_2} x_1^r x_2^s x_1 x_2 dx_1 dx_2 = \frac{2^{r+s+3}}{(r+2)(r+s+4)}.$$

The expressions for marginal moments are

$$\mu'_{r,0} = E\left(X_1^r\right) = \frac{2^{r+3}}{(r+2)(r+4)} \text{ and } \mu'_{0,s} = E\left(X_2^s\right) = \frac{2^{s+2}}{s+4}.$$

The means and variances of X_1 and X_2 are given as

$$\mu_1 = \frac{16}{15}; \mu_2 = \frac{8}{5};\ \sigma_{11} = \frac{44}{225}; \sigma_{22} = \frac{8}{75};$$

The covariance between X_1 and X_2 is

$$\sigma_{12} = E\{(X_1 - \mu_1)(X_2 - \mu_2)\}$$

$$= \frac{1}{2}\int_0^2\int_0^{x_2}\left(x_1 - \frac{16}{15}\right)\left(x_2 - \frac{8}{5}\right)x_1 x_2 dx_1 dx_2 = \frac{16}{225}.$$

The marginal distributions of X_1 and X_2 are

$$f_1(x_1) = \frac{1}{2}\int_{x_1}^2 x_1 x_2 dx_2 = \frac{1}{4}\left(4x_1 - x_1^3\right); 0 < x_1 < 2$$

$$f_2(x_2) = \frac{1}{2}\int_0^{x_2} x_1 x_2 dx_1 = \frac{x_2^3}{4}; 0 < x_2 < 2$$

Finally the conditional distribution of X_1 given $X_2 = x_2$ is

$$f(x_1|x_2) = \frac{2x_1}{x_2^2}; 0 < x_1 < x_2$$

The rth conditional moment is

$$E\left(X_1^r|x_2\right) = \int_0^{x_2} x_1^r \frac{2x_1}{x_2^2} dx_1 = \frac{2x_2^r}{r+2};$$

Consequently, the conditional mean and variance of X_1 given $X_2 = x_2$ are

$$\mu_{1|2} = \frac{2x_2}{3} \text{ and } \sigma_{11.2} = \frac{x_2^2}{18}.$$

Example 1.2: The joint density function of three random variables is

$$f(x_1, x_2, x_3) = \frac{4}{81}(x_1 + x_2 + x_3); 0 < x_1 < x_2 < x_3 < 3$$

Obtain the conditional distribution of (X_1, X_2) given $X_3 = x_3$. Also obtain the partial correlation coefficient between (X_1, X_2) given $X_3 = x_3$.

Solution: We first obtain the marginal distribution of X_3 as

$$f_3(x_3) = \frac{4}{81}\int_0^{x_3}\int_0^{x_2}(x_1 + x_2 + x_3)\,dx_1dx_2 = \frac{4}{81}x_3^3\ ;0 < x_3 < 3$$

The conditional distribution of (X_1, X_2) given $X_3 = x_3$ is therefore

$$f(x_1, x_2 | x_3) = \frac{f(x_1, x_2, x_3)}{f_3(x_3)} = \frac{x_1 + x_2 + x_3}{x_3^3}\ ;0 < x_1 < x_2 < x_3$$

The $(r,s)th$ conditional moment is

$$E(X_1^r X_2^s | x_3) = \int_0^{x_3}\int_0^{x_2} x_1^r x_2^s \frac{x_1 + x_2 + x_3}{x_3^3}dx_1dx_2$$

$$= \frac{x_3^{r+s}\{3r(r+s+4) + 5s + 12\}}{(r+s+3)(r+s+2)(r+2)(r+1)}.$$

The conditional means of X_1 given $X_3 = x_3$ and X_2 given $X_3 = x_3$ are

$$\mu_{1|3} = E(X_1 | x_3) = \frac{3x_3}{8} \text{ and } \mu_{2|3} = E(X_2 | x_3) = \frac{17x_3}{24}.$$

The conditional or partial variances of X_1 given $X_3 = x_3$ and X_2 given $X_3 = x_3$ are

$$\sigma_{11.3} = E(X_1^2 | x_3) - \{E(X_1 | x_3)\}^2 = \frac{19x_3^2}{320};$$

$$\sigma_{22.3} = E(X_2^2 | x_3) - \{E(X_2 | x_3)\}^2 = \frac{139x_3^2}{2880}.$$

The conditional or partial covariance between X_1 and X_2 given $X_3 = x_3$ is

$$\sigma_{12.3} = E(X_1X_2 | x_3) - \{E(X_1 | x_3)\}\{E(X_2 | x_3)\} = \frac{7x_3^2}{24}.$$

Finally, the partial correlation coefficient between X_1 and X_2 given $X_3 = x_3$ is

$$\rho_{12.3} = \frac{\sigma_{12.3}}{\sqrt{\sigma_{11.2}\sigma_{22.3}}} = \frac{280}{\sqrt{2641}}.$$

The multivariate theory is based upon the vectors and matrices and the study of matrix algebra is essentail for better understanding of the multivariate analyses. In the following we will, briefly, discuss some matrix algebra.

1.4 Some Matrix Algebra

Suppose we have a square (p x p) matrix **A** and a (p x 1) vector $\mathbf{x}$ which represent a point in p-dimensional *Euclidean space* V_p. In the following we will give some useful properties of vectors and matrices.

1.4.1 Vectors and their Properties

Some properties of vectors are listed below.

- The collection of all the vectors in V_p which are closed under addition and multiplication is called a vector space; that is if $\mathbf{x} \in V_p$ and $\mathbf{y} \in V_p$ then $\mathbf{x} + \mathbf{y} \in V_p$ and if α is a scaler and $\mathbf{x} \in V_p$ then $\alpha\mathbf{x} \in V_p$.
- A subset S of V_p is called vector subspace if S is itself a vector space.
- The set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\}$ are linearly dependent if there exist set of real numbers $\alpha_1, \alpha_2, \ldots, \alpha_k$; not all equal to zero; such that

$$\sum_{i=1}^{k} \alpha_i \mathbf{x}_i = 0, \qquad (1.4.1)$$

  otherwise the set of vectors are linearly independent.
- If **x** and **y** are two vectors of same order then the quantity

$$\mathbf{x}'\mathbf{y} = \sum_{i=1}^{k} x_i y_i; \qquad (1.4.2)$$

  is called *inner product* of the vectors.
- The quantity $\left(\mathbf{x}'\mathbf{x}\right)^{1/2}$ is called *norm* of the vector **x** and is denoted as $\|\mathbf{x}\|$.
- The *Euclidean distance* between two vectors is

$$\|\mathbf{x}-\mathbf{y}\| = \left[(\mathbf{x}-\mathbf{y})'(\mathbf{x}-\mathbf{y})\right]^{1/2}.$$

- The cosine of angle between two vectors is

$$\cos\theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|};\ 0 \le \theta \le 180^\circ. \qquad (1.4.3)$$

- Two vectors **x** and **y** are orthogonal if their inner product is zero. In this case the cosine of angle between **x** and **y** is zero.
- The *Minkowiski's norm* or *p-norm* of an $(n \times 1)$ vector **x** is given as

$$\|\mathbf{x}\|_p = \left\{ \sum_{i=1}^{n} |x_i|^p \right\}^{1/p} . \qquad (1.4.4)$$

1.4.2 Matrices and Their Properties

Some properties of matrices are given below.

- For two matrices **A** and **B** of same order the sum is defined as
  $\mathbf{C} = \mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$
- Product of a matrix **A** with a scaler α is defined as $\alpha\,\mathbf{A} = [\alpha a_{ij}]$
- The product of two matrices $\mathbf{A}_{n\times d}$ and $\mathbf{B}_{d\times m}$ is given as
  $$\mathbf{C}_{n\times m} = \mathbf{AB} = \left[c_{ij}\right] \text{for } c_{ij} = \sum_{k=1}^{d} a_{ik} b_{kj} .$$
- The trace of a (p x p) matrix **A** is sum of its diagonal elements; that is
  $$tr(\mathbf{A}) = \sum_{i=1}^{p} a_{ii} .$$
- The *squared matrix norm* is given as
  $$\|\mathbf{A}\|^2 = \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}^2 = tr\left(\mathbf{A}\mathbf{A}^/\right) = tr\left(\mathbf{A}^/\mathbf{A}\right); \qquad (1.4.5)$$
  and *Euclidean matrix norm* is simply $\|\mathbf{A}\|$.
- The distance between two matrices **A** and **B** is
  $$\|\mathbf{A}-\mathbf{B}\|^2 = tr\left[(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^/\right] = \sum_{i,j}\left(a_{ij}-b_{ij}\right)^2 . \qquad (1.4.6)$$
- The *vec operator* for an (n x p) matrix $\mathbf{A} = \left[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \cdots \quad \mathbf{a}_p\right]$ stacks the columns of **A** on each other so that to form a new (np x 1) vector **a** given as
  $$\mathbf{a} = vec(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \vdots \\ \mathbf{a}_p \end{bmatrix} .$$
  It is important to note that $tr\left(\mathbf{A}^/\mathbf{A}\right) = \|\mathbf{a}\|^2$.

- The *vech operator* on a symmetric (p x p) matrix **A** stacks diagonal below elements of the matrix. It has an order of $\left(\left\{p\left(p+1\right)/2\right\}\times 1\right)$.
- The *Direct* or *Kronecker* product of two matrices $\mathbf{A}_{n\times m}$ and $\mathbf{B}_{p\times q}$ is an $\left(np\times nq\right)$ matrix given as

$$\mathbf{A}\otimes\mathbf{B}=\begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & \cdots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & \cdots & a_{2m}\mathbf{B} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & \cdots & a_{nm}\mathbf{B} \end{bmatrix}. \quad (1.4.7)$$

  The Kronecker product always exist whether matrices are conformable for simple multiplication or not.
- The *Hadamard product* of two matrices of same order is defined as

$$\mathbf{A}\odot\mathbf{B}=\left[a_{ij}b_{ij}\right].$$

- The *Direct sum* of two matrices **A** and **B** is:

$$\mathbf{A}\oplus\mathbf{B}=\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}.$$

- Some trace properties associated with direct sum and direct products are given below

  (a) $tr\left(\mathbf{A}\otimes\mathbf{B}\right)=tr\left(\mathbf{A}\right)tr\left(\mathbf{B}\right)$

  (b) $tr\left(\mathbf{AB}\right)=\mathbf{1}_n^{/}\left(\mathbf{A}\odot\mathbf{B}\right)\mathbf{1}_m^{/}$

  (c) $tr\left\{\left(\mathbf{A}^{/}\odot\mathbf{B}^{/}\right)\mathbf{C}\right\}=tr\left\{\mathbf{A}^{/}\left(\mathbf{B}^{/}\odot\mathbf{C}\right)\right\}$

  (d) $tr\left[\oplus_{i=1}^{k}\mathbf{A}_i\right]=\sum_{i=1}^{k}tr\left(\mathbf{A}_i\right)$

  (e) $tr\left(\mathbf{A}^{/}\mathbf{B}\right)=\left\{vec\left(\mathbf{A}\right)\right\}^{/}\left\{vec\left(\mathbf{B}\right)\right\}$

  (f) $tr\left(\mathbf{ABC}\right)=vec\left(\mathbf{A}^{/}\right)\left(\mathbf{I}\otimes\mathbf{B}\right)vec\left(\mathbf{C}\right)$
- The rank of a square matrix **A** is equal to linearly independent rows in that matrix. In this case it is called *row rank*. Number of linearly independent columns of **A** is called *column rank*.
- For a matrix **A** we define its inverse as a matrix $\mathbf{A}^{-1}$ such that $\mathbf{AA}^{-1}=\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}$; where **I** is identity matrix. Matrix inverse has following properties

(a) $\left(\mathbf{AB}\right)^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

(b) $\left(\mathbf{A} \otimes \mathbf{B}\right)^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$

(c) $\left(\mathbf{I} + \mathbf{A}\right)^{-1} = \mathbf{A}\left(\mathbf{A} + \mathbf{I}\right)^{-1}$

(d) $\left(\mathbf{A}+\mathbf{B}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{A}+\mathbf{B}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}\mathbf{A}^{-1}$

so that $\mathbf{B}\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{A} = \left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}$.

(e) $\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1} = \left(\mathbf{I}+\mathbf{AB}^{-1}\right)^{-1}$

(f) For partitioned matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \mathbf{A}^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix};$$

where

$$\mathbf{B}_{11} = \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1}$$

$$\mathbf{B}_{12} = -\mathbf{B}_{11}\mathbf{B}_{12}\mathbf{A}_{22}^{-1}$$

$$\mathbf{B}_{21} = \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11}$$

$$\mathbf{B}_{22} = \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$$

provided all inverses exist.

- Determinant of a $(p \times p)$ matrix is a scaler quantity given as

$$\left|\mathbf{A}\right| = \sum (-1)^k a_{1i_1} a_{2i_2} \ldots a_{pi_p} \qquad (1.4.8)$$

where sum is taken over all $p!$ permutations of the elements of **A.**

- The (i, j) th Cofactor of matrix **A** is

$$C_{ij} = (-1)^{i+j} \left|\mathbf{M}_{ij}\right|;$$

where $\mathbf{M}_{ij}$ is a submatrix obtained from **A** by deleting i th row and j th column. In term of cofactors the determinant of **A** is given as $\left|\mathbf{A}\right| = \sum_j a_{ij} C_{ij} = \sum_i a_{ij} C_{ij}$.

- Some properties of determinant are

(a) $\left|\mathbf{A}\right| = \left|\mathbf{A}^{/}\right|$

(b) $\left|\mathbf{AB}\right| = \left|\mathbf{BA}\right|$

(c) For an orthogonal matrix $|\mathbf{A}| = \pm 1$.

(d) If $\mathbf{A}^2 = \mathbf{A}$ then $|\mathbf{A}| = 0$ or 1.

(e) For $\mathbf{A}_{p\times p}$ and $\mathbf{B}_{q\times q}$; $|\mathbf{A}\otimes\mathbf{B}| = |\mathbf{A}|^p\,|\mathbf{B}|^q$

(f) For partitioned matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

$$|\mathbf{A}| = |\mathbf{A}_{11}|\,|\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}|$$

$$= |\mathbf{A}_{22}|\,|\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|;$$

provided $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are invertible.

1.4.3 Eigenvalues and Eigenvectors

Let **A** be a $(p \times p)$ matrix. The quantities λ's are called eigen values or characteristic roots of the matrix **A** if **A** – λ**I** is singular. Hence the determinant of **A** – λ**I** is zero that is

$$|\mathbf{A} - \lambda\mathbf{I}| = 0, \tag{1.4.9}$$

and is a p degree polynomial in λ'*s* with eigen values $\lambda_1, \lambda_2, \ldots, \lambda_p$.

The vector $\mathbf{p}_i$ satisfying

$$(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{p}_i = 0 \text{ for } i = 1,2,\ldots,p, \tag{1.4.10}$$

is called eigen vector associated with λ_i.

The eignevalues and eigenvectors provide the basis for new matrices which are useful in studying further properties in certain multivariate cases. A popular technique is called *spectral decomposition* which has following properties

$$\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{\Lambda}; \mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda}; \mathbf{P}\mathbf{P}' = \mathbf{I} = \sum_i \mathbf{p}_i\mathbf{p}_i' \text{ and } \mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}' = \sum_i \lambda_i\mathbf{p}_i\mathbf{p}_i'; \tag{1.4.11}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with diagonal elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ and **P** is a $(p \times p)$ matrix with ith column $\mathbf{p}_i$. Rational powers of any matrices can be defined by using spectral decomposition. Specifically the square root matrix for a matrix **A** is defined as $\mathbf{A}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}$. Similarly square root matrix for inverse of a matrix is defined as $\mathbf{A}^{-1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}$. In general any rational power of matrix **A** is defined as $\mathbf{A}^{r/s} = \mathbf{P}\mathbf{\Lambda}^{r/s}$.

The generalized eignevalues and eigenvectors can also be defined as below.

Suppose we have $(p \times p)$ matrices **A** and **B**. The generalized eigenvalues of **A** in metric of **B** are given by solution of polynomial equation

$$|\mathbf{A} - \delta\mathbf{B}| = 0. \tag{1.4.12}$$

The generalized eigenvectors **A** in metric of **B** are non-trivial solution of:

$$(\mathbf{A} - \delta_i\mathbf{B})\mathbf{q}_i = 0. \tag{1.4.13}$$

If we define matrix **Q** with ith column $\mathbf{q}_i$ and $\mathbf{\Delta}$ as diagonal matrix with diagonal element δ_i; then the generalized spectral decomposition has following properties

$$\mathbf{\Delta} = \mathbf{Q}'\mathbf{AQ} \text{ and } \mathbf{Q}'\mathbf{BQ} = \mathbf{I};$$

$$\mathbf{A} = \left(\mathbf{Q}'\right)^{-1}\mathbf{\Delta}\mathbf{Q}^{-1} \text{ and } \left(\mathbf{Q}'\right)^{-1}\mathbf{Q}^{-1} = \mathbf{B},$$

$$\mathbf{A} = \sum_i \delta_i \mathbf{q}_i\mathbf{q}_i' \text{and} \mathbf{B} = \sum_i \mathbf{q}_i\mathbf{q}_i'.$$

The generalized eigenvalues and eigenvectors are useful in multivariate statistical models.

Example 1.3: For the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 5 & -2 \\ 0 & -2 & 1 \end{bmatrix},$$

Obtain eigenvalues and eigenvectors.

Solution: The characteristic equation is

$$|\mathbf{A} - \lambda\mathbf{I}| = 0$$

or $\lambda^3 - 7\lambda^2 + 6\lambda = 0$

Hence the eigenvalues are $\lambda_1 = 6$, $\lambda_2 = 1$ and $\lambda_3 = 0$. The eigenvectors are

$$\mathbf{p}_1 = \begin{bmatrix} -1 \\ -5 \\ 2 \end{bmatrix}; \mathbf{p}_2 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \text{ and } \mathbf{p}_3 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}.$$

The properties given in spectral decomposition can be easily checked.

1.5 Multivariate Descriptive Measures

The descriptive measures are useful to study certain properties of the variables, for example the mean of a variable provides a measure of central tendency of the data. The variance provides a measure of dispersion of a univariate data from its mean. In multivariate analysis the data on several variables is available and that data can be used to obtain the descriptive measures for each variable individually and in pairs and are called *multivariate descriptive measures*. We will discuss these multivariate descriptive measures in the following.

1.5.1 The Mean Vector

The mean vector for a random vector **x** is a $(p \times 1)$ vector who's ith entry is expected value of ith random variable X_i. It is denoted by $\boldsymbol{\mu}$. More specifically

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ \vdots \\ E(X_2) \end{bmatrix} = E(\mathbf{x}) \cdot$$

The sample mean vector for a $(p \times n)$ data matrix is computed as

$$\bar{\mathbf{x}} = \frac{1}{n}\mathbf{X1} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j, \tag{1.5.1}$$

where $\mathbf{1}$ is an $(n \times 1)$ vector of 1'*s* and $\mathbf{x}_j$ is jth column of **X**.

1.5.2 The Covariance Matrix

In univariate analysis, the variance provides information about variability in the data from the mean. In case of a univariate distribution the variance provide information about variability of the random variable from its expected value. In multivariate analysis we deal with several random variables simultaneously or we have random sample on several variables and we can compute variance of each variable. Additionally, we can compute the covariance for all possible pairs of the variables. These variances and covariances can be collected in the form of a matrix. This matrix is known as the *covariance matrix*. In case of the joint distribution of p random variables, the covariance matrix is defined as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \cdots & \sigma_{1h} & \cdots & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \cdots & \sigma_{2h} & \cdots & \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{i1} & \sigma_{i2} & \cdots & \cdots & \sigma_{ih} & \cdots & \cdots & \sigma_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \cdots & \sigma_{ph} & \cdots & \cdots & \sigma_{pp} \end{bmatrix} = \left[\sigma_{ih}\right].$$

The *covariance matrix* for the sample looks like $\boldsymbol{\Sigma}$ but the population variances and covariances are replaced with their sample counterparts. The covariance matrix shows variability in the multivariate data. The covariance matrix for a random vector is obtained as

$$\boldsymbol{\Sigma} = E\left\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'\right\}, \tag{1.5.2}$$

where $\mathbf{x}'$ is transpose of the vector $\mathbf{x}$. The covariance matrix is a symmetric positive definite matrix such that $|\boldsymbol{\Sigma}| > 0$. The quantity $tr(\boldsymbol{\Sigma})$ is called *total variance* and $|\boldsymbol{\Sigma}|$ is called the *generalized variance*.

The sample covariance matrix is computed as

$$\mathbf{S} = \frac{1}{n-1}\left(\mathbf{X}\mathbf{X}' - n\bar{\mathbf{x}}\bar{\mathbf{x}}'\right)$$
$$= \frac{1}{n-1}\sum\nolimits_{j=1}^{n}\left(\mathbf{x}_j - \bar{\mathbf{x}}\right)\left(\mathbf{x}_j - \bar{\mathbf{x}}\right)' = \frac{1}{n-1}\left(\sum\nolimits_{j=1}^{n}\mathbf{x}_j\mathbf{x}_j' - n\bar{\mathbf{x}}\bar{\mathbf{x}}'\right). \tag{1.5.3}$$

The sample mean vector and sample covariance matrix are unbiased estimators of their population counterparts.

1.5.3 The Correlation Matrix

The correlation coefficient is an important measure to study the streangth of interdependence between two random variables or between variables of a random sample drawn from a bivariate population. In multivariate analysis we have information on several random variables and hence the correlation coefficients can be computed for all possible pairs. These correlation coefficients can be collected in the form of a ($p \times p$) matrix $\boldsymbol{\rho}$ which is given as

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1i} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2i} & \cdots & \rho_{2p} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \rho_{i1} & \rho_{i2} & \cdots & 1 & \cdots & \rho_{ip} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pi} & \cdots & 1 \end{bmatrix} = \left[\rho_{ij}\right];$$

and is called the *population correlation matrix*. The correlation matrix can be computed from the covariance matrix by using

$$\boldsymbol{\rho} = \mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{-1/2}; \tag{1.5.4}$$

where $\mathbf{V} = diag\left[\sigma_{ii}\right]$. The sample counterpart of the correlation matrix can be analogously defined. The correlation matrix has the property that $0 \le |\boldsymbol{\rho}|^2 \le 1$ and if $|\boldsymbol{\rho}|^2$ is 1 then this indicates that the variables are completely uncorrelated. The sample correlation matrix can be defined analogously.

1.5.4 Generating Functions for Random Vectors and Matrices

The generating functions are useful in studying the properties of distribution of a random variable. The moment generating function is one of the generating function which is helpful in obtaining moments of a random variable. The moment generating function for multivariate case is an extension of univariate moment generating function and can be used to obtain the moments of a random vector. The moment generating function for a random vector is defined below.

Suppose we have a random vector $\mathbf{x}$ with joint density function $f\left(x_1, x_2, ..., x_p\right)$ or compactly $f\left(\mathbf{x}\right)$ then the moment generating function for random vector $\mathbf{x}$ is defined as

$$M_{\mathbf{x}}\left(\mathbf{t}\right) = E\left(e^{\mathbf{t}'\mathbf{x}}\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{\mathbf{t}'\mathbf{x}} f\left(\mathbf{x}\right) dx_1 \cdots dx_p, \tag{1.5.5}$$

Provided that it exist. The moment generating function is useful in obtaining moments of a random vector.

Another important generating function for a random vector is the characteristic function which is defined as

$$\phi_{\mathbf{x}}(\mathbf{t}) = E\left(e^{i\mathbf{t}^{/}\mathbf{x}}\right) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i\mathbf{t}^{/}\mathbf{x}} f(\mathbf{x})\,dx_1\cdots dx_p, \quad (1.5.6)$$

and this function always exists. The characteristic function can be used to obtain the density function by using the inversion theorem. The characteristic function is a unique function as compared with the moment generating function.

The cumulants are useful measures to describe the properties of a random vacriable. The cumulant generating function for a random vector provides basis to study its properties. The cumulant generating function for a random vector is

$$\psi_{\mathbf{x}}(\mathbf{t}) = \ln\{\phi_{\mathbf{x}}(\mathbf{t})\}. \quad (1.5.7)$$

The moment generating function can be extended to the case of random matrices. The moment generating function of a random matrix **X** is defined as

$$M_{\mathbf{X}}(\mathbf{T}) = E\left[\exp\left\{tr\left(\mathbf{T}^{/}\mathbf{X}\right)\right\}\right] = E\left\{e^{vec^{/}\mathbf{T}vec\mathbf{X}}\right\}. \quad (1.5.8)$$

The characteristic function for a random matrix is similarly defined.

The moment generating function defined above is used to obtain the raw moments of a random variable. The centeral moments are more useful as compared with the raw moments in studying the properties of a distribution. Although, the raw moments can be used to compute the central moments but it is sometime more convenient to compute the central moments directly. The central moments of a distribution are easily computed by using the *central moment generating function,* which for a random vector **x** is defined as

$$M_{\mathbf{x}-\mu}(\mathbf{t}) = E\left\{e^{\mathbf{t}^{/}(\mathbf{x}-\mu)}\right\} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{\mathbf{t}^{/}(\mathbf{x}-\mu)} f(\mathbf{x})\,dx_1\cdots dx_p$$
$$= e^{-\mathbf{t}^{/}\mu}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{\mathbf{t}^{/}\mathbf{x}} f(\mathbf{x})\,dx_1\cdots dx_p = e^{-\mathbf{t}^{/}\mu} M_{\mathbf{x}}(\mathbf{t}). \quad (1.5.9)$$

In similar way the central moment generating function for a random matrix **X** is given as

$$M_{\mathbf{X}-E(\mathbf{X})}(\mathbf{T}) = \exp\left[tr\left\{\mathbf{T}^{/}E(\mathbf{X})\right\}\right] M_{\mathbf{X}}(\mathbf{T}). \quad (1.5.10)$$

The moment generating function and characteristic function are useful in obtaining moments of random vectors and random matrices.