CHAPTER 5

MULTIVARIATE POPULATIONS

5.1 INTRODUCTION

In the following chapters we will be dealing with a variety of problems concerning multivariate populations. The purpose of this chapter is to provide an introduction to the notions and basic inference methods concerning parameters of a multivariate normal population. We will discuss the issues that arise in making simultaneous inferences concerning means of correlated variables. We will also address the problem of performing regressions on a number of correlated variables. The problem of comparing a number of multivariate populations will be addressed in Chapter 6, in which the multivariate regression procedures will play a key role.

We encounter multivariate populations whenever we have to deal with two or more measurements on some quantities that are correlated with each other. A sample taken from any of the multivariate populations would then consist of a vector of independent observations. This random vector may be a number of distinct response variables observed for each of the experimental units comprising the sample. They could also be repeated measurements of one variable observed at distinct time points. In short, we are now dealing with

(Generalized Inference in Repeated Measures, Edition 2). By (Weerahandi)

a vector of possibly correlated responses or variables from a certain population as opposed to a single variable representing the population. The vector of observations may deal with the same characteristic, related characteristics, or completely different characteristics concerning the experimental unit. For example, the level of education and income of an individual may be two characteristics we need to deal with. In another application, the correlated variables that one is interested might be some test scores of students in a number of college courses.

The purpose of this chapter is to outline some basic notions and inference methods in multivariate data analysis. In particular, various inferences concerning the mean vector of a multivariate normal population will be discussed. The treatment of this chapter will cover both the classical procedures and new procedures in this context including methods for computing exact confidence regions for the parameters of a multivariate regression model. The problem of making inferences about the difference in the mean vectors of two multivariate normal populations including exact solutions to the Multivariate Behrens−Fisher problem will be addressed in Chapter 6.

5.1.1 Notations

We will use bold face letters such as $\mathbf{X}$ and $\mathbf{Y}$ to denote a multivariate random vectors representing a population of interest. If the random vector $\mathbf{X}$ has p components, then its components are denoted in alternative forms as

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = (X_1 \ X_2 \ \cdots \ X_p)' .$$

The *mean vector* of a population will be denoted by lowercase Greek letters such as $\boldsymbol{\mu}$ and the covariance matrix by uppercase Greek letters such as $\boldsymbol{\Sigma}$. If $\boldsymbol{\mu}$ represents the mean vector of a random vector $\mathbf{X}$, it is defined as

$$\boldsymbol{\mu} = \mathbf{E}(\mathbf{X}) = \begin{pmatrix} EX_1 \\ EX_2 \\ \vdots \\ EX_p \end{pmatrix} . \tag{5.1}$$

The *covariance matrix* of $\mathbf{X}$ is defined as

$$\boldsymbol{\Sigma} = \mathbf{Var}(\mathbf{X}) = \mathbf{E}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})', \tag{5.2}$$

$$= \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_p) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_p) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \mathrm{Cov}(X_p, X_2) & \cdots & \mathrm{Var}(X_p) \end{pmatrix}, \tag{5.3}$$

respectively. Of course the covariance matrix is symmetric, as evident from

$$\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = E((X_i - E(X_i))(X_j - E(X_j)).$$

The *characteristic function* of the random vector $\mathbf{X}$ is defined as

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbf{E}\mathbf{e}^{i\mathbf{t}'\mathbf{X}},$$

where $i = \sqrt{-1}$ is the complex number representing the square root of -1 and $\mathbf{t}$ is any $p \times 1$ vector of real numbers. The characteristic function of a random vector is not only useful in deducing the moments of the random vector, but also particularly useful in deriving the distribution of linear combinations of random vectors. The latter is possible because the characteristic function uniquely identifies the distribution of the random vector. Some useful properties of the characteristic function are as follows.

i. $\mathbf{E}(\mathbf{X}) = \frac{1}{i}\frac{d\phi_{\mathbf{X}}(\mathbf{t})}{d\mathbf{t}}|_{\mathbf{t}=\mathbf{0}}$.

ii. $\mathbf{Var}(\mathbf{X}) = \frac{d^2\phi_{\mathbf{X}}(\mathbf{t})}{d\mathbf{t}d\mathbf{t}'}|_{\mathbf{t}=\mathbf{0}} - \mathbf{E}(\mathbf{X})\mathbf{E}(\mathbf{X}')$.

iii. If $\mathbf{X}$ and $\mathbf{Y}$ are independent random vectors of same dimension, then the characteristic function of their sum is equal to the product of their characteristic functions; i.e.,

$$\phi_{\mathbf{X}+\mathbf{Y}}(\mathbf{t}) = \phi_{\mathbf{X}}(\mathbf{t})\phi_{\mathbf{Y}}(\mathbf{t}). \tag{5.4}$$

iv. If the random vector $\mathbf{X}$ and a $\mathbf{t}$ vector of constants are partitioned with the same dimension as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{t} = \begin{pmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{pmatrix},$$

then,
(a) the characteristic function of the marginal distribution of $\mathbf{X}_1$ is found as $\phi_{\mathbf{X}_1}(\mathbf{t}_1) = \phi_{\mathbf{X}}(\mathbf{t}_1, \mathbf{0})$,
(b) the two random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if

$$\phi_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{t}_1, \mathbf{t}_2) = \phi_{\mathbf{X}_1}(\mathbf{t}_1)\phi_{\mathbf{X}_2}(\mathbf{t}_2). \tag{5.5}$$

As usual, a sample of size n taken from a multivariate population represented by a random vector $\mathbf{X}$ is denoted by $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. Unless otherwise mentioned, we will assume that these are independent random observations.

5.2 MULTIVARIATE NORMAL POPULATIONS

In this book we are concerned only with normal populations. If some original distribution was not multivariate normal, we assume that it has already been transformed into one that is multivariate normal. The definition of a multivariate normal population is given below, which, in particular, implies that each component of the random vector representing the population is normal; that is, $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \ldots, p$, where $\mu_i = E(X_i)$ and $\sigma_i^2 = E(X_i - \mu_i)^2$. Of course in the current context these components are possibly correlated and hence we have the need to express their joint distribution. Muirhead (1982) and Anderson (1984), in particular, provide a comprehensive discussion of the multivariate normal distribution and its properties. Here we outline some main results that we will need in the applications undertaken in this book.

Let $\mathbf{X}$ be a random vector representing a multivariate population. Suppose its mean vector is $\boldsymbol{\mu}$ and its covariance matrix is $\boldsymbol{\Sigma}$. Assume that $\boldsymbol{\Sigma}$ is positive definite. Then, the joint probability density of a multivariate normal vector $\mathbf{X}$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}. \tag{5.6}$$

That the p-dimensional random vector $\mathbf{X}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix is $\boldsymbol{\Sigma}$ is denoted as

$$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{5.7}$$

The characteristic function of $\mathbf{X}$ is given by

$$\phi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}'\boldsymbol{\mu} - \mathbf{t}'\Sigma\mathbf{t}/2}. \tag{5.8}$$

A few properties of the multivariate normal distribution that will prove to be very useful in applications undertaken in this book are
(i) marginals of multivariate normal distribution are also normal,
(ii) linear transformations of a multivariate normal random vector is also multivariate normal, and
(iii) the sum of two p-variate normal random vectors is also a p-variate normal random vector.

Property (i) means that, the distribution of any subvector of the random matrix $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ also has normal distribution with corresponding subvector of mean and the subcovariance matrix. That is, if

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix},$$

is a partition of $\mathbf{X}$, then the distribution of the subvector $\mathbf{X}_1$ is given by

$$\mathbf{X}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}), \tag{5.9}$$

where $q(< p)$ is the dimension of $\mathbf{X}_1$ vector and, the $q \times 1$ mean vector $\boldsymbol{\mu}_1$ and the $q \times q$ covariance matrix $\boldsymbol{\Sigma}_{11}$ are obtained from the obvious partitions of the same dimension, namely

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

To provide the precise formula for the second property, let $\mathbf{A}$ be a $q \times p$ full rank matrix of constants and $\mathbf{b}$ be a $q \times 1$ vector of constants, $q \leq p$. Then, the distribution of the random vector $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ is given by

$$\mathbf{Y} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'). \tag{5.10}$$

Finally, Property (iii) means that if $\mathbf{X}_1 \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathbf{X}_2 \sim N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ are independent random vectors, then

$$\mathbf{X}_1 + \mathbf{X}_2 \sim N_p(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2).$$

We can employ Property (ii) to standardize a multivariate normal vector, namely,

$$\text{if } \mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ then } \mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, \mathbf{I}_p), \tag{5.11}$$

where $\boldsymbol{\Sigma}^{-1/2}$ is a positive definite square root of matrix of $\boldsymbol{\Sigma}$; if $\boldsymbol{\Sigma} = \mathbf{P}'\mathbf{D}_\lambda\mathbf{P}$ is a diagonalization of $\boldsymbol{\Sigma}$, then the required square root is defined as

$$\boldsymbol{\Sigma}^{-1/2} = \mathbf{P}'\mathbf{D}_{\sqrt{\lambda}}\mathbf{P},$$

where $\mathbf{P}$ is a $p \times p$ orthogonal matrix formed by eigenvectors of $\boldsymbol{\Sigma}$, $\mathbf{D}_\lambda$ is the diagonal matrix formed by eigenvalues of $\boldsymbol{\Sigma}$, and $\mathbf{D}_{\sqrt{\lambda}}$ is the diagonal matrix formed by positive square roots of eigenvalues of $\boldsymbol{\Sigma}$. From the second and the third properties we can deduce another result that will prove to be very useful in sampling from a normal population. If $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2, \cdots, n$ is a set of independent random vectors and a_i, $i = 1, 2, \cdots, n$ is a set of constants, then

$$\sum a_i \mathbf{X}_i \sim N_p(\sum a_i \boldsymbol{\mu}_i, \sum a_i^2 \boldsymbol{\Sigma}_i). \tag{5.12}$$

The above properties can be established by using the properties of the characteristic function (see Exercises 5.1 and 5.2) or direct manipulation of the probability density function. For example, to prove Property (ii), we can first find the characteristic function of the random vector $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ as

$$\begin{aligned}
\phi_{\mathbf{Y}}(\mathbf{t}) &= \mathbf{E}\mathbf{e}^{i\mathbf{t}'\mathbf{Y}} \\
&= \mathbf{E}\mathbf{e}^{i\mathbf{t}'(\mathbf{AX}+\mathbf{b})} \\
&= \mathbf{e}^{i\mathbf{t}'\mathbf{b}}\mathbf{E}\mathbf{e}^{i(\mathbf{t}'\mathbf{A})\mathbf{X}} \\
&= \mathbf{e}^{i\mathbf{t}'\mathbf{b}} e^{i(\mathbf{t}'\mathbf{A})\boldsymbol{\mu} - (\mathbf{t}'\mathbf{A})\Sigma(\mathbf{t}'\mathbf{A})'/2} \\
&= e^{i\mathbf{t}'(\mathbf{A}\boldsymbol{\mu}+\mathbf{b}) - \mathbf{t}'\mathbf{A}\Sigma\mathbf{A}'\mathbf{t}/2}.
\end{aligned}$$

But this is the characteristic function of the multivariate normal distribution with the mean vector $\mathbf{A}\boldsymbol{\mu} + \mathbf{b}$ and the covariance matrix $\mathbf{A}\Sigma\mathbf{A}'$, thus proving the property.

5.2.1 Wishart distribution

Another important distribution closely related to the multivariate normal distribution is the Wishart distribution. In fact in the order of importance and usefulness in multivariate data analysis, it ranks second only to the multivariate normal distribution. The Wishart distribution is a multivariate generalization of the univariate gamma distribution and often arise in making inferences about parameters of the multivariate normal distribution. In fact the definition of the Wishart distribution was motivated by this need that arised in sampling from a multivariate normal population.

If $\mathbf{Z}_i \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, $i = 1, 2, \ldots, n$, are independently and identically distributed multivariate random variables, then the random matrix

$$\mathbf{Q} = \sum_{i=1}^{n} \mathbf{Z}_i \mathbf{Z}_i{}' \tag{5.13}$$

is said to have a p-dimensional Wishart distribution with n degrees of freedom and parameter matrix $\mathbf{\Sigma}$ and is denoted as $\mathbf{Q} \sim W_p(n, \mathbf{\Sigma})$. The joint probability density function of $\mathbf{Q}$ is given by

$$f(\mathbf{Q}) = \frac{c|\mathbf{Q}|^{(n-p-1)/2}}{|\Sigma|^{n/2}} e^{-\frac{1}{2} tr \Sigma^{-1} \mathbf{Q}} \quad \text{for } \mathbf{Q} > \mathbf{0}, \tag{5.14}$$

where c is a constant and by $\mathbf{Q} > \mathbf{0}$ we mean that $\mathbf{Q}$ is positive definite. The characteristic function of the Wishart distribution is

$$\phi_{\mathbf{Y}}(\mathbf{T}) = |\mathbf{I}_p - 2i\mathbf{\Sigma}\mathbf{T}|^{-n/2}, \tag{5.15}$$

where $\mathbf{T}$ is any symmetric matrix of real numbers.

The reader is referred to Rao (1973), Muirhead (1982), and Anderson (1984) for various properties and derivations of the Wishart distribution. Displayed below are some properties that we will be using in this following chapters. The derivations of these properties are straightforward from the definition of the Wishart distribution and the properties of the multivariate normal distributions.

(i). If $\mathbf{Q} \sim W_p(n, \mathbf{\Sigma})$ and if $\mathbf{A}$ is a $q \times p$ constant matrix with $q \leq p$, then

$$\mathbf{A}\mathbf{Q}\mathbf{A}' \sim W_q(n, \mathbf{A}\mathbf{\Sigma}\mathbf{A}'), \tag{5.16}$$

provided that $\mathbf{A}$ is of full rank.

(ii). If $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are two independent Wishart matrices distributed as $\mathbf{Q}_1 \sim W_p(m, \mathbf{\Sigma})$ and $\mathbf{Q}_2 \sim W_p(n, \mathbf{\Sigma})$, then the distribution of their sum is also Wishart. More specifically,

$$\mathbf{Q}_1 + \mathbf{Q}_2 \sim W_p(m+n, \mathbf{\Sigma}). \tag{5.17}$$

For Wishart distributions, it is also true that if $\mathbf{Q}_1 \sim W_p(m, \mathbf{\Sigma})$ and $\mathbf{Q}_1 + \mathbf{Q}_2 \sim W_p(m+n, \mathbf{\Sigma})$, then $\mathbf{Q}_2 \sim W_p(n, \mathbf{\Sigma})$, provided that $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are independent, an implication of (5.17) that easily proved using the characteristics function of the Wishart distribution.

(iii). If $\mathbf{Q} \sim W_p(n, \mathbf{\Sigma})$ is partitioned as

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix},$$

where is $\mathbf{Q}_{11}$is a $q \times q$ matrix with $q \leq p$, then the distribution of $\mathbf{Q}_{11}$ is given by

$$\mathbf{Q}_{11} \sim W_q(n, \mathbf{\Sigma}_{11}), \tag{5.18}$$

where $\mathbf{\Sigma}_{11}$ is the $q \times q$ matrix obtained by partitioning the matrix Σ the same way as $\mathbf{Q}$:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}.$$

5.3 INFERENCES ABOUT THE MEAN VECTOR

In the following sections we use the symbols $\mathbf{X}$ and $\mathbf{Y}$ to denote quantities that are different from the above sections. Let us now use $\mathbf{Y}$ to denote a vector of p responses of interest from a multivariate population and suppose that data are available from n subjects or experimental units. Assume that $\mathbf{Y}$ has a multivariate normal distribution with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\mathbf{\Sigma}$. Let $\mathbf{Y}_i$ be the $p \times 1$ vector of responses from ith subject. Assume that

$$\mathbf{Y}_i \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma}), \quad i = 1, 2, \ldots, n, \tag{5.19}$$

are mutually independent random vectors and that $n > p$. The main purpose of this section is to develop procedures for making inferences about the parameter vector $\boldsymbol{\mu}$, the parameter of primary interest. Except for point estimation, the problem of making inferences about the covariance matrix $\mathbf{\Sigma}$ will be deferred until the next section.

In view of the definitions (5.7) and (5.8) of the parameters $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$, define the sample mean vector and the sample covariance matrix as

$$\overline{\mathbf{Y}} = \sum_{i=1}^{n} \mathbf{Y}_i / n \tag{5.20}$$

and

$$\mathbf{S} = \frac{\sum_{i=1}^{n} (\mathbf{Y}_i - \overline{\mathbf{Y}})(\mathbf{Y}_i - \overline{\mathbf{Y}})'}{n}, \tag{5.21}$$

respectively. In fact, these are maximum likelihood estimates (MLEs) of parameters. That the sample mean vector is indeed the MLE of the mean parameter vector $\boldsymbol{\mu}$ is clear when the likelihood function is expressed as

$$\begin{aligned} L(\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\sum(y_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(y_i-\boldsymbol{\mu})} \\ &= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\sum(y_i-\overline{\mathbf{y}})'\boldsymbol{\Sigma}^{-1}(y_i-\overline{\mathbf{y}})+n(\overline{\mathbf{y}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{y}}-\boldsymbol{\mu})} \\ &= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{n}{2}tr(\boldsymbol{\Sigma}^{-1}[\mathbf{S}+(\overline{\mathbf{y}}-\boldsymbol{\mu})(\overline{\mathbf{y}}-\boldsymbol{\mu})'])}. \end{aligned} \tag{5.22}$$

The MLE of the covariance matrix $\boldsymbol{\Sigma}$ can also be established by differentiating the log likelihood function

$$\begin{aligned} l(\boldsymbol{\mu},\boldsymbol{\Lambda}) &= \log(L(\boldsymbol{\mu},\boldsymbol{\Lambda})) \\ &= -\frac{np}{2}\log(2\pi)+\frac{n}{2}\log(|\boldsymbol{\Lambda}|)-\frac{n}{2}\text{tr}(\boldsymbol{\Lambda}\mathbf{A}) \end{aligned}$$

and equating it to $\mathbf{0}$, where $\mathbf{A}=\mathbf{S}+(\overline{\mathbf{y}}-\boldsymbol{\mu})(\overline{\mathbf{y}}-\boldsymbol{\mu})'$ and $\boldsymbol{\Lambda}=\boldsymbol{\Sigma}^{-1}$; the reader is referred to Press (1982) for formulas for vector and matrix differentiations. Differentiating $l(\boldsymbol{\mu},\boldsymbol{\Lambda})$ with respect to $\boldsymbol{\Lambda}$ we get

$$\frac{dl(\boldsymbol{\mu},\boldsymbol{\Lambda})}{d\boldsymbol{\Lambda}}=\frac{n}{2}\boldsymbol{\Lambda}^{-1}-\frac{n\mathbf{A}}{2}.$$

Since $\mathbf{A}$ reduces to $\mathbf{S}$ at the MLE of $\boldsymbol{\mu}$, the MLE of $\boldsymbol{\Lambda}$ is $\mathbf{S}$, thus implying the desired result. Two other important results that follow from (5.22) are that $(\overline{\mathbf{Y}},\mathbf{S})$ are sufficient statistics for unknown parameters $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ and that $\overline{\mathbf{Y}}$ and $\mathbf{S}$ are independently distributed.

While $\overline{\mathbf{Y}}$ is an unbiased estimate of $\boldsymbol{\mu}$, the unbiased estimate of $\boldsymbol{\Sigma}$ based on $\mathbf{S}$ is

$$\widehat{\boldsymbol{\Sigma}}=n\mathbf{S}/(n-1)=\frac{\sum_{i=1}^{n}(\mathbf{Y}_i-\overline{\mathbf{Y}})(\mathbf{Y}_i-\overline{\mathbf{Y}})'}{n-1}. \tag{5.23}$$

The former is obvious from the expected values of the individual components of $\overline{\mathbf{Y}}$ and the latter can be easily proved by using the orthogonal decomposition

$$\sum_{i=1}^{n}(\mathbf{Y}_i-\boldsymbol{\mu})(\mathbf{Y}_i-\boldsymbol{\mu})'=\sum_{i=1}^{n}(\mathbf{Y}_i-\overline{\mathbf{Y}})(\mathbf{Y}_i-\overline{\mathbf{Y}})'+n(\overline{\mathbf{Y}}-\boldsymbol{\mu})(\overline{\mathbf{Y}}-\boldsymbol{\mu})' \tag{5.24}$$

and taking expected values of term by term to obtain

$$n\boldsymbol{\Sigma}=\mathbf{E}\sum_{i=1}^{n}(\mathbf{Y}_i-\overline{\mathbf{Y}})(\mathbf{Y}_i-\overline{\mathbf{Y}})'+n\boldsymbol{\Sigma}/n.$$

The orthogonal decomposition (5.24) also implies that the sample mean vector and the sample covariance matrix are independently distributed.

From (5.12) we get $\sum \mathbf{Y}_i \sim N_p(n\boldsymbol{\mu}, n\boldsymbol{\Sigma})$ and in turn

$$\overline{\mathbf{Y}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}), \tag{5.25}$$

It follows from the definition (5.21), (5.24), and Property (ii) of Wishart distributions that the distribution of $\mathbf{S}$ is given by

$$n\mathbf{S} = \sum_{i=1}^{n}(\mathbf{Y_i} - \overline{\mathbf{Y}})(\mathbf{Y_i} - \overline{\mathbf{Y}})' \sim W_p(n-1, \boldsymbol{\Sigma}). \tag{5.26}$$

For more rigorous proofs of the above results, the reader is referred to Rao (1973), Muirhead (1982), or Anderson(1984).

Example 5.1. Study of Mathematics and Physics scores

Table 5.1 presents a set of standardized scores of a random sample of college students. Let us first estimate the population parameters. Let $\boldsymbol{\mu}$ be the 2×1 vector comprised of the student population score in Mathematics (X) and that in Physics (Y). Let $\boldsymbol{\Sigma}$ be the 2×2 matrix comprised of variances of scores in the two subjects and the covariance between them. In this application, the sample size and the dimension of the multivariate population are $n = 10$ and $p = 2$, respectively. The unbiased estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ obtained by applying equations (5.21) and (5.23) are as follows:

$$\widehat{\boldsymbol{\mu}} = \begin{pmatrix} 62.6 \\ 66.6 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 164.9 & 36.0 \\ 36.0 & 70.0 \end{pmatrix}$$

The mean score in Physics seems to be greater than that in Mathematics. Later in this chapter we will test whether this difference is statistically significant or it is an artifact of the small sample size. Notice also that the variances of the scores are fairly high. The *correlation coefficient* between Physics and Mathematics scores can be computed using the covariance and variances as

$$\begin{aligned} \widehat{\rho} &= \frac{36.0}{\sqrt{164.9 * 70.0}} \\ &= 0.335 \end{aligned}$$

5.3.1 Hypothesis Testing

Now consider the problem of testing the mean vector based on a random sample $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ from a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Suppose the problem is to test the null hypothesis

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \tag{5.27}$$

Table 5.1 Standardized test scores

Mathematics	**Physics**
72.8	69.9
46.0	68.9
59.2	58.4
66.7	78.2
84.2	63.9
50.4	54.6
49.6	66.5
77.9	71.6
63.9	77.2
55.1	56.8

against the alternative hypothesis $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. If the covariance matrix were known, then the inferences about the mean vector $\boldsymbol{\mu}$, including the problem of testing the above hypothesis, can be based on the chi-squared statistic

$$V = n(\overline{\mathbf{Y}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{Y}} - \boldsymbol{\mu}) \sim \chi^2_p. \tag{5.28}$$

That the random variable V has a chi-squared distribution follows from 5.11, because V can be expressed in terms of a set of standard normal random variates as

$$V = \mathbf{Z}\mathbf{Z}' = \sum_{i=1}^{p} Z_i^2,$$

where $\mathbf{Z} = \sqrt{n}\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, \mathbf{I}_p)$ and Z_i's are the components of $\mathbf{Z}$.

Typically $\boldsymbol{\Sigma}$ is unknown and so one instead uses

$$T^2 = n(\overline{\mathbf{Y}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{Y}} - \boldsymbol{\mu}_0) \tag{5.29}$$

as the test statistic, which is commonly known as the *Hotelling's T^2 statistic* with p and $n-1$ degrees of freedom. The distribution of the Hotelling's T^2 statistic is related to a noncentral F distribution as

$$F(\delta) = \frac{(n-p)}{p(n-1)}T^2 = \frac{n-p}{p}(\overline{\mathbf{Y}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{Y}} - \boldsymbol{\mu}_0) \sim F_{p,n-p}(\delta), \tag{5.30}$$

where $\delta = (\boldsymbol{\mu} - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)$ is the noncentrality parameter. For a proof of this result, the reader is referred to Rao (1973) or Muirhead (1982). Under the null hypothesis (5.27), the noncentral F distribution appearing in (5.30) reduces to a central F distribution.

Since the right-tail probability of the T^2 statistic increases for deviations from the null hypothesis, the right-tail probability serves as an unbiased test of H_0, The p-value of the test can be computed as

$$\begin{aligned} p &= \Pr\left(F \geq \frac{n-p}{p}(\overline{\mathbf{y}}-\boldsymbol{\mu}_0)'\mathbf{s}^{-1}(\overline{\mathbf{y}}-\boldsymbol{\mu}_0)\right) \\ &= 1 - H_{p,n-p}(\frac{n-p}{p}(\overline{\mathbf{y}}-\boldsymbol{\mu}_0)'\mathbf{s}^{-1}(\overline{\mathbf{y}}-\boldsymbol{\mu}_0), \end{aligned} \tag{5.31}$$

where $\overline{\mathbf{y}}$ is the observed value of $\overline{\mathbf{Y}}$, $\mathbf{s}$ is the observed value of $\mathbf{S}$, and $H_{p,n-p}$ is the cumulative distribution function (cdf) of the F distribution with p and $n-p$ degrees of freedom. The null hypothesis H_0 is rejected if the observed value of p is too small; at fixed-level α, H_0 is rejected if $p \leq \alpha$. It can be shown that the test given by (5.31) is also the likelihood ratio test of H_0.

Example 5.2. Study of Mathematics and Physics scores (continued)

Consider again the data reported in Table 5.1. Now suppose we are interested in testing the hypothesis

$$H_0 : \boldsymbol{\mu} = \begin{pmatrix} 65 \\ 70 \end{pmatrix}.$$

The Hotelling T^2 statistic computed using (5.29) is

$$\begin{aligned} T^2 &= 10 \begin{pmatrix} -2.4 & -3.4 \end{pmatrix} \begin{pmatrix} 164.9 & 36.0 \\ 36.0 & 70.0 \end{pmatrix}^{-1} \begin{pmatrix} -2.4 \\ -3.4 \end{pmatrix} \\ &= 1.75. \end{aligned}$$

Hence, the p-value for testing the hypothesis can be computed as

$$\begin{aligned} p &= 1 - H_{2,8}\left(\frac{8}{2 \times 9} \times 1.75\right) \\ &= 1 - .509 \\ &= 0.491. \end{aligned}$$

This p-value does not support the rejection of the null hypothesis.

5.3.2 Confidence Regions

Confidence regions corresponding to the above unbiased test are easily derived from the quantiles of the central F distribution,

$$F = \frac{n-p}{p}(\overline{\mathbf{Y}}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{Y}}-\boldsymbol{\mu}) \sim F_{p,n-p}. \tag{5.32}$$

The $100\gamma\%$ confidence ellipsoid given by this approach can be written as

$$\frac{n-p}{p}(\overline{\mathbf{y}}-\boldsymbol{\mu})'\mathbf{s}^{-1}(\overline{\mathbf{y}}-\boldsymbol{\mu}) \leq F_{p,n-p}(\gamma) \tag{5.33}$$

or as

$$(\overline{\mathbf{y}}-\boldsymbol{\mu})'\widehat{\boldsymbol{\Sigma}}^{-1}(\overline{\mathbf{y}}-\boldsymbol{\mu})\leq\left(\frac{p(n-1)}{n(n-p)}\right)F_{p,n-p}(\gamma),$$

where, for convenience, the observed value of $\widehat{\boldsymbol{\Sigma}}$ is denoted by the same notation, $F_{p,n-p}(\gamma)$ is the γth quantile of the F distribution with p and $n-p$ degrees of freedom.

Example 5.3. Study of Mathematics and Physics scores (continued)

Consider again the data set reported in Table 5.1. We can now use formula (5.33) to construct a confidence ellipse for the mean vector $\boldsymbol{\mu}=\begin{pmatrix}\mu_x & \mu_y\end{pmatrix}'$. The 95% confidence ellipse given by the formula is

$$0.0069\Delta\mu_x^2+0.0169\Delta\mu_y^2-0.0074\Delta\mu_x\Delta\mu_y<1.003,$$

where

$$\Delta\mu_x=\mu_x-62.6, \quad \text{and} \quad \Delta\mu_y=\mu_y-66.6.$$

Note, in particular that the vector of values hypothesized in Example 5.2, namely $\boldsymbol{\mu}=\begin{pmatrix}65 & 70\end{pmatrix}'$ falls well within this ellipse, because $0.0069\,(2.4)^2+0.0169\,(3.4)^2-0.0074\times(2.4)\,(3.4)=0.175$ is much less than 1.003.

5.4 INFERENCES ABOUT LINEAR FUNCTIONS OF $\boldsymbol{\mu}$

Now suppose we need to test a certain linear combination of components of the mean vector $\boldsymbol{\mu}$. To be specific, consider a linear hypotheses of the form

$$H_0:\mathbf{A}\boldsymbol{\mu}=\mathbf{b}, \tag{5.34}$$

where $\mathbf{A}$ is a $q\times p$ matrix of rank q with known constants, $\mathbf{b}$ is a $q\times 1$ vector of some hypothesized constants, and $q\leq p$. Procedures for testing the hypothesis can be deduced from the foregoing results. It follows from (5.20) that

$$\mathbf{A}\overline{\mathbf{Y}}\sim N_q(\mathbf{A}\boldsymbol{\mu},\frac{1}{n}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}') \tag{5.35}$$

and hence the corresponding Hotelling's T^2 statistic becomes

$$T^2=n(\mathbf{A}\overline{\mathbf{y}}-\mathbf{b})'(\mathbf{A}\widehat{\boldsymbol{\Sigma}}\mathbf{A}')^{-1}(\mathbf{A}\overline{\mathbf{y}}-\mathbf{b}). \tag{5.36}$$

Now testing of the hypothesis (5.34) can be carried out using its distribution given by

$$F=\frac{n-q}{q(n-1)}T^2=\frac{n-q}{q}(\mathbf{A}\overline{\mathbf{Y}}-\mathbf{b})'(\mathbf{A}\mathbf{S}\mathbf{A}')^{-1}(\mathbf{A}\overline{\mathbf{Y}}-\mathbf{b})\sim F_{q,n-q} \tag{5.37}$$

when H_0 is true. It is now evident that the p-value appropriate for testing H_0 can be computed as

$$p = 1 - H_{q,n-q}\left(\frac{n-q}{q}(\mathbf{A}\overline{\mathbf{y}} - \mathbf{b})'(\mathbf{A}\mathbf{s}\mathbf{A}')^{-1}(\mathbf{A}\overline{\mathbf{y}} - \mathbf{b})\right) \tag{5.38}$$

and that the null hypothesis is rejected for small values of the p-value. Note, in particular, that when $p = 2$ and $\mathbf{A} = (\ 1 \quad -1\ )$, this test reduces to the classical paired t-test. More generally, this procedure could be utilized to test the equality of all components of the mean vector $-$ i.e., the equality of the means of p correlated populations

$$H_0 : \mu_1 = \mu_1 = \cdots = \mu_p. \tag{5.39}$$

To derive a test for this hypothesis we can simply define

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix}$$

set $\mathbf{b} = \mathbf{0}$, and then compute the p-value using (5.38) with $q = p - 1$.

5.4.1 Simultaneous confidence regions

Now consider the problem of making confidence statements about one or more sets of parameters based on the components of the mean vector $\boldsymbol{\mu}$. The usual $100\gamma\%$ confidence ellipsoid for the parameter vector $\boldsymbol{\eta} = \mathbf{A}\boldsymbol{\mu}$ immediately follows from (5.37) as

$$\frac{n-q}{q}(\mathbf{A}\overline{\mathbf{y}} - \boldsymbol{\eta})'(\mathbf{A}\mathbf{s}\mathbf{A}')^{-1}(\mathbf{A}\overline{\mathbf{y}} - \boldsymbol{\eta}) \leq F_{q,n-q}(\gamma). \tag{5.40}$$

In particular, $100\gamma\%$ confidence intervals for one prespecified linear combination of the form $\eta = \mathbf{a}'\boldsymbol{\mu}$ can be deduced from (5.40) or derived from the fact that

$$\mathbf{a}'\overline{\mathbf{Y}} \sim N_1(\eta, \frac{1}{n}\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}) \quad \text{and} \quad (n-1)\frac{\mathbf{a}'\widehat{\boldsymbol{\Sigma}}\mathbf{a}}{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} \sim \chi^2_{n-1}, \tag{5.41}$$

where $\mathbf{a}$ is a $p \times 1$ vector of known constants. From results concerning univariate normal distributions, it is now obvious that the $100\gamma\%$ equal-tail confidence interval for η can be obtained as

$$\mathbf{a}'\overline{\mathbf{y}} \pm t_{n-1}((1+\gamma)/2)\left(\frac{\mathbf{a}'\widehat{\boldsymbol{\Sigma}}\mathbf{a}}{n}\right)^{1/2}, \tag{5.42}$$

where $t_{n-1}(\nu)$ is the νth quantile of the Student's t distribution with $n-1$ degrees of freedom.

If there were r prespecified linear combinations, $\mathbf{a}_i'\boldsymbol{\mu}$ of interest, simultaneous confidence intervals for all parameters could be obtained by applying the Bonferroni method. The conservative simultaneous intervals of level at least $100\gamma\%$ that follow directly from (5.42) are

$$\mathbf{a}_i'\overline{\mathbf{y}} \pm t_{n-1}(1-\alpha/2r)\left(\frac{\mathbf{a}_i'\widehat{\boldsymbol{\Sigma}}\mathbf{a}_i}{n}\right)^{1/2}, \quad i = 1, 2, \ldots, r, \tag{5.43}$$

where $\alpha = 1 - \gamma$.

Scheffe-type simultaneous confidence intervals for all possible linear functions of $\boldsymbol{\eta}$ or linear functions of $\boldsymbol{\mu}$ can be obtained using (5.40) with $\mathbf{A}' = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_q]$ formed by all or first $q \leq \min(r, p)$ components of the first linearly independent (without loss of generality) q vectors thus making $rank(\mathbf{A}) = q$. Letting $\mathbf{V} = ((n-q)/q)\mathbf{AsA}'$ and $\boldsymbol{\delta} = (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})$, it follows from the Scheffe inequality that

$$\begin{aligned} \gamma &= \Pr\left[\boldsymbol{\delta}'\mathbf{V}^{-1}\boldsymbol{\delta} \leq F_{q,n-q}(\gamma)\right] \\ &= \Pr\left(\frac{(\mathbf{c}'\delta)^2}{\mathbf{c}'\mathbf{V}\mathbf{c}} \leq F_{q,n-q}(\gamma) \text{ for all } \mathbf{c}\right) \\ &= \Pr(\mathbf{c}'\eta \in \mathbf{c}'\widehat{\eta} \pm (\mathbf{c}'\mathbf{V}\mathbf{c}F_{q,n-q}(\gamma))^{1/2} \text{ for all } \mathbf{c}). \end{aligned} \tag{5.44}$$

Since the set of intervals

$$\mathbf{a}_i'\overline{\mathbf{y}} \pm k_\gamma \left(\frac{\mathbf{a}_i\widehat{\boldsymbol{\Sigma}}\mathbf{a}_i'}{n}\right)^{1/2}, \quad i = 1, 2, \ldots, r \tag{5.45}$$

is a subset of the above intervals, the confidence level of simultaneous intervals computed using (5.45) is at least γ, where

$$k_\gamma = [q(n-1)F_{q,n-q}(\gamma)/(n-q)]^{1/2}.$$

If the desired linear combinations are prespecified, then the Bonferroni intervals are usually much shorter and hence preferred over the Scheffe intervals. If they are prespecified and yet r is much larger than q, then the Scheffe intervals are preferred. If the linear combinations are not pre-specified, then the confidence level of the Bonferroni intervals is not valid. In this case the formula (5.45) is applied with $q = p$ if simultaneous confidence intervals for any number of arbitrary linear combinations of means to be constructed, and it is applied with $q = p - 1$ if confidence intervals for all linear contrasts are desired, such as all possible pairwise differences of means. In particular, in constructing simultaneous confidence intervals for the individual means we define $\mathbf{a}_i$ to be a vector of zeros except for the ith element which is 1. These confidence intervals as well as the tests discussed above can be computed using most statistical packages such as SAS, SPlus, SPSS, and XPro.

Example 5.4. Comparing cork borings thickness

Consider the data shown in Table 5.2, which was studied by Rao (1948) to test whether the thickness of cork borings (in cm) on trees was the same in the north, east, south, and west directions.

Table 5.2 Thickness of cork borings

N	**E**	**W**	**S**
72	66	76	77
56	57	64	58
32	32	35	36
39	39	31	27
37	40	31	25
32	30	34	28
54	46	60	52
91	79	100	75
79	65	70	61
78	55	67	60
39	35	34	37
60	50	67	54
39	36	39	31
43	37	39	50
60	53	66	63
41	29	36	38
30	35	34	26
42	43	31	25
33	29	27	36
63	45	74	63
47	51	52	43
56	68	47	50
81	80	68	58
46	38	37	38
32	30	30	32
35	37	48	39
50	34	37	40
48	54	57	43

In this problem $p = 4$, $n = 28$, and the sample mean and the sample covariances are

$$\overline{\mathbf{y}}=\begin{pmatrix} 50.54 \\ 46.18 \\ 49.68 \\ 45.18 \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}=\begin{pmatrix} 290.4 & 223.8 & 288.4 & 226.3 \\ 223.8 & 219.9 & 229.1 & 171.4 \\ 288.4 & 229.1 & 350.0 & 259.5 \\ 226.3 & 171.4 & 259.5 & 226.0 \end{pmatrix},$$

the unbiased estimated of the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, respectively. Let $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3, \delta_4)' = \boldsymbol{\mu} - \overline{\mathbf{y}}$. In this case, in constructing 95%

confidence regions we get $F_{p,n-p}(\gamma) = F_{4,24}(.95) = 2.776$, $p(n-1)/n(n-p) = 0.1607$, and the inverse of the sample covariance matrix is

$$\widehat{\mathbf{\Sigma}}^{-1} = \begin{pmatrix} 0.0321 & -0.0158 & -0.0079 & -0.0111 \\ -0.0158 & 0.0221 & -0.0050 & 0.0048 \\ -0.0079 & -0.0050 & 0.02667 & -0.0190 \\ -0.0111 & 0.0048 & -0.0190 & 0.0337 \end{pmatrix}.$$

Therefore, the 95% confidence ellipsoid given by (5.33) for the mean vector can be written as

$$\begin{aligned} &0.0321\delta_1^2 + 0.0227\delta_2^2 + 0.0267\delta_3^2 + 0.0337\delta_4^2 \\ &+2\delta_1\delta_2(-0.0158) + 2\delta_1\delta_3(-0.0078) + 2\delta_1\delta_4(-0.0111) \\ &+2\delta_2\delta_3(-0.0050) + 2\delta_2\delta_4(0.0048) + 2\delta_3\delta_4(-0.0190) < 0.4462 \end{aligned}$$

To illustrate the use of hypotheses testing procedure, consider the particular hypothesis $H_0 : \boldsymbol{\mu} = \begin{pmatrix} 40 & 40 & 40 & 40 \end{pmatrix}$. The Hotelling's T^2 statistic in this case is 20.742, leading to an F-statistic of 6.40186 with 3 and 25 degrees of freedom, and a p-value of 0.00659. This means that there is strong evidence to rejects the null hypothesis. The value of the Hotelling's T^2 statistic for testing the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is 326.181, leading to an F-statistic of 72.4847 with 4 and 24 degrees of freedom, and a p-value of 0.00228. This p-value also suggests the rejection of the hypothesis. The table below shows the 95% simultaneous intervals for individual means computed using the formulas (5.43) and (5.45).

Scheffe Intervals		
Mean	Lower Bound	Upper Bound
μ_1	39.15	61.92
μ_2	36.27	56.08
μ_3	37.18	62.18
μ_4	35.14	55.22
Bonferroni Intervals		
Mean	Lower Bound	Upper Bound
μ_1	41.92	59.15
μ_2	38.68	53.68
μ_3	40.22	59.14
μ_4	37.58	52.78

In constructing the Scheffe intervals for individual means, the appropriate degrees of freedom for the F distribution are 4 and 24. In constructing intervals for any number of contrasts using the Scheffe method, in view of remarks made concerning (5.44), we could use the F distribution with 3 and 25 degrees of freedom. In particular, the 95% simultaneous intervals for all possible 6 pairs of mean differences are shown below along with Bonferroni intervals obtained by setting $r = 6$. Note that the Bonferroni intervals are shorter in both cases. The size of the Bonferroni intervals would be larger if we were interested in some additional contrasts, whereas the length of the Scheffe intervals would remain unchanged regardless of the number of contrasts of interest. From these intervals, Rao (1948) made some interesting observations. For example, the intervals suggest that the mean thickness of cork borings in the direction of South is significantly different from that of East.

Scheffe Intervals		
Contrast	Lower Bound	Upper Bound
$\mu_1 - \mu_2$	-0.306	9.021
$\mu_2 - \mu_3$	-9.721	2.721
$\mu_3 - \mu_4$	0.061	8.939
$\mu_4 - \mu_1$	-10.06	-0.655
$\mu_1 - \mu_3$	-3.832	5.547
$\mu_2 - \mu_4$	-4.976	6.976
Bonferroni Intervals		
Contrast	Lower Bound	Upper Bound
$\mu_1 - \mu_2$	0.093	8.622
$\mu_2 - \mu_3$	-9.189	2.189
$\mu_3 - \mu_4$	0.441	8.559
$\mu_4 - \mu_1$	-9.657	-1.058
$\mu_1 - \mu_3$	-3.431	5.146
$\mu_2 - \mu_4$	-4.465	6.465

5.5 MULTIVARIATE REGRESSION

In model (5.19) we assumed that the mean vector of $\mathbf{Y}_i$, the $p \times 1$ vector of responses, were the same for all n observations. When this is not true due to some uncontrolled variables, we could still solve the problem by taking a regression approach as in the univariate case, provided that some of the uncontrolled variables are observable. We do so by fitting a separate univariate regression for each component of the $\mathbf{Y}$ vector. Since the response variable of these regressions are correlated, the inferences about underlying parameters can be made by taking a multivariate approach. In short, multivariate regression is a multivariate extension of the univariate regression when we are interested in a number of correlated dependent variables. As we will see in

Chapter 6, multivariate regressions play an important role when we need to compare a number of treatment groups based on some data that have been collected according to some design. In the literature on regressions and their applications, the vector $\mathbf{Y}$ is referred to as the dependent random vector, or the vector of responses.

Let $\mathbf{x}_i$ be a $q \times 1$ vector of covariates available for describing some of the variation in $\mathbf{Y}_i$. Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be set of independent observations of the random vector $\mathbf{Y}$ while the covariates are possibly changing. Assuming a regression model for each component of $\mathbf{Y}_i$ we have

$$Y_{ij} = \mathbf{x}_i' \boldsymbol{\beta}_j + \varepsilon_{ij}, \quad i = 1, \ldots n \text{ and } j = 1, \ldots, p. \tag{5.46}$$

The univariate regression model on jth component of the response variable implied by (5.46) is

$$\mathbf{Y}_j = \mathbf{X}\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j, \tag{5.47}$$

where $\mathbf{Y}_j$ is a $n \times 1$ vector formed by observations on the jth component of the response variable, $\mathbf{X}$ is a $n \times q$ matrix formed by all observed values of covariates and $\boldsymbol{\beta}_j$ is a $q \times 1$ vector of parameters. In applications involving various designed experiments that we will undertake in the following chapters, the $\mathbf{X}$ matrix is often referred to as the *design matrix.* If Y_{ij} values are used to form $\mathbf{Y}_i'$, a $1 \times n$ raw vector, we get another form of the model as

$$\mathbf{Y}_i' = \mathbf{x}_i' \mathbf{B} + \boldsymbol{\varepsilon}_i, \tag{5.48}$$

where $\mathbf{B}$ is the $q \times p$ matrix of parameters formed by all $\boldsymbol{\beta}_j$s, which is to be estimated based on the independent observations $\mathbf{Y}_i$, $i = 1, \ldots, n$. Some authors use this form of the model instead of the form (5.47) to introduce the multivariate regression model. By piling column vectors in (5.47) or row vectors in (5.48) we can express the two alternative forms of the model in a compact form as

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\varepsilon}, \tag{5.49}$$

where $\mathbf{Y}$ is a $n \times p$ matrix formed by observation from all dependent variables. The matrix $\mathbf{X}$ in this model is referred to as the matrix of covariates, explanatory variables, or as the design matrix in various applications.

To make various inferences about the parameter matrix $\mathbf{B}$ beyond the point estimation, we make the usual normality assumption that $\boldsymbol{\varepsilon}_i \sim N_p(0, \boldsymbol{\Sigma})$. We also assume that $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ are independently distributed. These assumptions imply that

$$Vec(\boldsymbol{\varepsilon}) \sim N_{np}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n) \text{ and } Vec(\mathbf{Y}) \sim N_{np}(Vec(\mathbf{XB}), \boldsymbol{\Sigma} \otimes \mathbf{I}_n), \tag{5.50}$$

where by $Vec(\boldsymbol{\varepsilon})$ we mean the $np \times 1$ vector obtained by piling components of $\boldsymbol{\varepsilon}$ all in one column.

5.5.1 Estimation of regression parameters

As in the case of a common mean vector we discussed above, the regression parameter matrix $\mathbf{B}$ can be estimated by applying a series of univariate regressions using the form of the model (5.47). In other words, the covariance matrix does not enter into the point estimation of components of $\mathbf{B}$. As will become clear later, however, the covariance matrix is important in interval estimation and in hypothesis testing. From the univariate regression formula for the estimate of $\boldsymbol{\beta}_j$, namely $\widehat{\boldsymbol{\beta}}_j = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_j$, we can then find the estimate of $\mathbf{B}$ by piling the columns of estimated parameters as

$$\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \tag{5.51}$$

This is in fact the MLE as well as the unbiased estimate of $\mathbf{B}$ based on sufficient statistics. To show that MLE of $\mathbf{B}$ is $\widehat{\mathbf{B}}$, we can use (5.50) to express the likelihood function as

$$\begin{aligned} L(\mathbf{B}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} f(\mathbf{y}_i) \\ &= c(\Sigma)\exp(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{Y}'_i - \mathbf{x}'_i\mathbf{B})\boldsymbol{\Sigma}^{-1}(\mathbf{Y}'_i - \mathbf{x}'_i\mathbf{B})') \\ &= c(\Sigma)\exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{X}\mathbf{B})'(\mathbf{Y}-\mathbf{X}\mathbf{B}))\right), \end{aligned} \tag{5.52}$$

where $c(\Sigma) = (2\pi)^{-np/2}|\Sigma|^{-n/2}$. Now by decomposing $\mathbf{B}$ dependent factors in (5.52) as

$$(\mathbf{Y}-\mathbf{X}\mathbf{B})'(\mathbf{Y}-\mathbf{X}\mathbf{B}) = \mathbf{Y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y} + (\mathbf{B}-\widehat{\mathbf{B}})'(\mathbf{X}'\mathbf{X})(\mathbf{B}-\widehat{\mathbf{B}}), \tag{5.53}$$

we can see that the likelihood function given by (5.52) is maximized when $\mathbf{B} = \widehat{\mathbf{B}}$, and that the MLE does not depend on the covariance matrix $\boldsymbol{\Sigma}$. By differentiating (5.53) with respect to $\boldsymbol{\Sigma}$, as we did in Section 5.3, it is also seen (see Exercise 5.9) that

$$\widetilde{\boldsymbol{\Sigma}} = \frac{1}{n}(\mathbf{Y}-\mathbf{X}\widehat{\mathbf{B}})'(\mathbf{Y}-\mathbf{X}\widehat{\mathbf{B}}) = \frac{1}{n}\mathbf{Y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y} \tag{5.54}$$

is the MLE of $\boldsymbol{\Sigma}$. The unbiased estimate of $\boldsymbol{\Sigma}$ based on the same statistics is

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n-q}\mathbf{Y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}. \tag{5.55}$$

Some important results that follow from the decomposition of (5.53) and from the properties of the multivariate normal and Wishart distributions are:

(i) $\widehat{\mathbf{B}}$ and $\widehat{\boldsymbol{\Sigma}}$ are sufficient statistics for ($\mathbf{B}$, $\boldsymbol{\Sigma}$).

(i) $\widehat{\mathbf{B}}$ and $\widehat{\boldsymbol{\Sigma}}$ are independently distributed.

(iii) The distribution of $\widehat{\mathbf{B}}$ is given by

$$\text{Vec}(\widehat{\mathbf{B}}) \sim \mathbf{N}_{pq}(\text{Vec}(\mathbf{B}), \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}). \tag{5.56}$$

(iv) The distribution of $\widehat{\boldsymbol{\Sigma}}$ is given by

$$(n-q)\widehat{\boldsymbol{\Sigma}} \sim W_p(n-q, \boldsymbol{\Sigma}). \tag{5.57}$$

5.5.2 Confidence regions

As in univariate regressions, it is easy to construct confidence regions on a single component $\boldsymbol{\beta}_j$ of $\mathbf{B}$ based on the result $\widehat{\boldsymbol{\beta}_j} \sim \mathbf{N}_q(\boldsymbol{\beta}_j, \sigma_{jj}^2(\mathbf{X}'\mathbf{X})^{-1})$ and the fact that the σ_{jj}^2 can be tackled by its estimate, which is related to the chi-squared distribution with $n-q$ degrees of freedom that follow from equation (5.57). Confidence regions on the whole matrix of parameters $\mathbf{B}$ or its submatrices are constructed based on the results,

$$\begin{aligned} \mathbf{E} &= (n-p)\widehat{\boldsymbol{\Sigma}} \sim W_p(n-q, \boldsymbol{\Sigma}), & (5.58) \\ \mathbf{H} &= (\widehat{\mathbf{B}} - \mathbf{B})'(\mathbf{X}'\mathbf{X})(\widehat{\mathbf{B}} - \mathbf{B}) \sim W_p(q, \boldsymbol{\Sigma}). & (5.59) \end{aligned}$$

That the second random matrix $\mathbf{H}$ also has a Wishart distribution is an implication of (5.56). In general, if $\mathbf{E} \sim W_d(e, \boldsymbol{\Sigma})$ and $\mathbf{H} \sim W_d(h, \boldsymbol{\Sigma})$, the following random variable, which arises from the likelihood ratio method,

$$U = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}|} = \frac{1}{|\mathbf{I}_d + \mathbf{H}\mathbf{E}^{-1}|},$$

is said to have a *U distribution* with d, h, and e degrees of freedom and is denoted as

$$U \sim U_{d,h,e}\ . \tag{5.60}$$

In terms of the U distribution, one can obtain a joint $100\gamma\%$ confidence region for $\mathbf{B}$ as

$$\{\mathbf{B} \mid |\mathbf{I}_p + \frac{1}{n}(\widehat{\mathbf{B}} - \mathbf{B})'(\mathbf{X}'\mathbf{X})(\widehat{\mathbf{B}} - \mathbf{B})\widehat{\boldsymbol{\Sigma}}^{-1}| \le \kappa_\gamma\}, \tag{5.61}$$

where κ_γ is the $1-\gamma$th quantile of the U distribution with p, q and $n-q$ degrees of freedom. The probabilities of the U distribution are easily evaluated using a set of independent univariate Beta random variates. The representation due to Anderson expresses the U random variable as a product of beta random variates as

$$U = B_1 B_2 \cdots B_d\ , \text{ where } B_k \sim \text{Beta}(\frac{e-k+1}{2}, \frac{h}{2}), \tag{5.62}$$

provided that $e \geq d$. The required condition $e \geq d$ is usually the case when we have adequate data for parameter estimation. When the dimension d of the underlying Wishart distributions or h parameter is small, there are simpler representations in terms of the F distribution:

$$\text{When } d = 1, \quad \frac{1-U}{U}\frac{e}{h} \sim F_{h.e}\ . \tag{5.63}$$

$$\text{When } d = 2 \text{ and } h \geq 2, \ \frac{1-U^{1/2}}{U^{1/2}}\frac{e-1}{h} \sim F_{2h.2(e-1)}. \tag{5.64}$$

$$\text{When } h = 1, \ \frac{1-U}{U}\frac{e+1-d}{d} \sim F_{d.e+1-d}. \tag{5.65}$$

$$\text{When } h = 2 \text{ and } d \geq 2, \ \frac{1-U^{1/2}}{U^{1/2}}\frac{e+1-d}{h} \sim F_{2d.2(e+1-d)}. \tag{5.66}$$

When d is large, perhaps the best way to compute probabilities and quantiles of the U distribution is using a large number of random numbers generated from each of the independent beta distributions in (5.62). The computations are carried out in the following steps.

i. Generate a large N number, say $N = 10000$, random numbers from each of the Beta random variables $B_1, B_2, \ldots, B_d$.

ii. Compute N random numbers from the U random variable using the formula $U = B_1 B_2 \cdots B_d$.

iii. Estimate probabilities of the form $\Pr(U \leq u)$ required in hypothesis testing by the fraction of U random numbers that are less than or equal to the value of u.

iv. Sort the U random numbers in ascending order.

v. Estimate the quantiles required in constructing confidence regions using the corresponding value of the ordered U data; for example, when $N = 10000$, estimate the 95th quantile of the U distribution by the 9500th value of the ordered data set.

On the basis of the two Wishart statistics $\mathbf{E}$ and $\mathbf{H}$, it is also possible to construct confidence regions for general linear combinations and contrasts of the form, $\Phi = \mathbf{CBD}$, and test hypotheses of the form

$$H_0 : \mathbf{CBD} = \mathbf{0},$$

where $\mathbf{C}$ and $\mathbf{D}$ are two matrices of constants. Appropriate choice of $\mathbf{C}$ and $\mathbf{D}$ allows us to perform comparisons of contrasts of row-wise or column-wise parameters of $\mathbf{B}$. Let us differ this type of analysis as well as illustrations until next chapter where the importance of such hypotheses becomes further clear.

Exercises

5.1 Let $\mathbf{X} \sim N_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a multivariate normal vector. Consider the partition of $\mathbf{X}$,

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix},$$

so that $\mathbf{X}_1$ is of dimension $p \times 1$, $p < r$. Using Property (iv) of characteristic functions, show that the marginal distribution of $\mathbf{X}_1$ is given by

$$\mathbf{X}_1 \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}),$$

where p is the dimension of $\mathbf{X}_1$ vector and, the $p \times 1$ mean vector $\boldsymbol{\mu}_1$ and the $p \times p$ covariance matrix $\boldsymbol{\Sigma}_{11}$ are obtained from the partitions of the same dimension as,

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

5.2 Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ and $\mathbf{Y} \sim N_p(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x)$ be independent random vectors of same dimension. Using Property (iii) of characteristic functions, show that $\mathbf{X} + \mathbf{Y} \sim N_p(\boldsymbol{\mu}_x + \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y)$. Hence deduce that if $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2, \ldots, n$ is a set of independent random vectors and a_i, $i = 1, 2, \ldots, n$ is a set of constants, then

$$\sum a_i \mathbf{X}_i \sim N_p(\sum a_i \boldsymbol{\mu}_i, \sum a_i^2 \boldsymbol{\Sigma}_i).$$

5.3 Using the properties of the characteristic function of the Wishart distribution, show that if $\mathbf{Q} \sim W_p(n, \boldsymbol{\Sigma})$ and if $\mathbf{A}$ is a $q \times p$ is a constant matrix with $q \leq p$, then

$$\mathbf{AQA}' \sim W_q(n, \mathbf{A\Sigma A}').$$

Also deduce the above result from properties of the multivariate normal distribution and the definition of the Wishart distribution given by (5.13).

5.4 If $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are two independent Wishart matrices distributed as

$$\mathbf{Q}_1 \sim W_p(m, \boldsymbol{\Sigma}) \quad \text{and} \quad \mathbf{Q}_2 \sim W_p(n, \boldsymbol{\Sigma}),$$

show (i) by using the characteristic function of the Wishart distribution and (ii) by applying the definition of the Wishart distribution given by (5.13) and using the properties of multivariate normal distribution that the distribution of their sum is given by

$$\mathbf{Q}_1 + \mathbf{Q}_2 \sim W_p(m + n, \boldsymbol{\Sigma}).$$

5.5 Let $\mathbf{Y}_j \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; $j = 1, \ldots, n$ be a random sample from a multivariate normal population. By differentiating the log likelihood function given by (5.21), derive the MLE of the mean vector $\boldsymbol{\mu}$.

5.6 Let $\mathbf{Y}_j \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; $j = 1, \ldots, n$ be a random sample from a multivariate normal population. Show that the MLE of $\boldsymbol{\Sigma}$ can also be written as

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Y}_i\mathbf{Y}' - \overline{\mathbf{Y}}\overline{\mathbf{Y}}'.$$

5.7 By expanding the identity

$$\sum_{i=1}^{n}(\mathbf{Y}_i - \boldsymbol{\mu})(\mathbf{Y}_i - \boldsymbol{\mu})' = \sum_{i=1}^{n}\left((\mathbf{Y}_i - \overline{\mathbf{Y}}) + (\overline{\mathbf{Y}} - \boldsymbol{\mu})\right)\left((\mathbf{Y}_i - \overline{\mathbf{Y}}) + (\overline{\mathbf{Y}} - \boldsymbol{\mu})\right)',$$

show that

$$\sum_{i=1}^{n}(\mathbf{Y}_i - \boldsymbol{\mu})(\mathbf{Y}_i - \boldsymbol{\mu})' = \sum_{i=1}^{n}(\mathbf{Y}_i - \overline{\mathbf{Y}})(\mathbf{Y}_i - \overline{\mathbf{Y}})' + n(\overline{\mathbf{Y}} - \boldsymbol{\mu})(\overline{\mathbf{Y}} - \boldsymbol{\mu})'$$

is an orthogonal decomposition. Also argue why this result implies that the sample mean vector and the sample covariance matrix are independently distributed.

5.8 Let $\mathbf{Y}_j \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; $j = 1, \ldots, n$ be a random sample from a multivariate normal population. Using the characteristic function of the Wishart distribution and the above results to prove that

$$\sum_{i=1}^{n}(\mathbf{Y}_i - \overline{\mathbf{Y}})(\mathbf{Y}_i - \overline{\mathbf{Y}})' \sim W_p(n-1, \boldsymbol{\Sigma}).$$

5.9 Consider the problem of testing the hypothesis

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$$

based on the random sample in Exercise 5.1. Show that the likelihood ratio test is the same as the test given by (5.31).

5.10 Consider the multivariate regression model

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_{np}(0, \boldsymbol{\Sigma} \otimes \mathbf{I}_n).$$

By differentiating the log likelihood function

$$\ln(L) = c + \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}tr\left(\boldsymbol{\Sigma}^{-1}(n\widetilde{\boldsymbol{\Sigma}} + (\mathbf{B} - \widehat{\mathbf{B}})'(\mathbf{X}'\mathbf{X})(\mathbf{B} - \widehat{\mathbf{B}}))\right),$$

with respect to $\mathbf{B}$ and $\boldsymbol{\Sigma}$, show that

$$\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad \text{and} \quad \widetilde{\boldsymbol{\Sigma}} = \frac{1}{n}\mathbf{Y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}$$

are their maximum likelihood estimates.

Issue 1	Issue 2	Issue 3
178	128	156
146	135	139
125	127	132
167	156	163
142	148	152
98	102	104
130	124	132
148	131	139
164	140	155
155	132	156

5.11 Consider the sample of multivariate data shown in table below. The data represents the weekly sales of three issues of a certain magazine at a sample of supermarkets.

Assuming that the underlying population is multivariate normal,

(a) construct a 95% confidence region for the vector of means $\boldsymbol{\mu} = (\mu_{1,}\mu_{2,}\mu_3)$,
(b) construct 95% simultaneous confidence intervals for $\mu_1 - \mu_2$ and $\mu_2 - \mu_3$,
(c) test the hypothesis that $H_0 : \mu_1 > \mu_2$.

5.12 Consider the data in Table 5.1. Compute p-values and carry out tests of each of following hypotheses:

(a) $H_0 : \mu_1 + \mu_2 + \mu_3 + \mu_4 \leq 195$,
(b) $H_0 : (\mu_1 + \mu_2) = (\mu_3 + \mu_4)$,
(c) $H_0 : \begin{pmatrix} \mu_1 + \mu_2 & \mu_3 + \mu_4 \end{pmatrix} \geq \begin{pmatrix} 100 & 100 \end{pmatrix}$.

5.13 Consider again the data in Table 5.1. Construct 95% confidence regions for each of the following quantities:

(a) sum of all four means,
(b) parameter, $(\mu_1 + \mu_2) - (\mu_3 + \mu_4)$,
(c) parameter vector, $\Theta = \begin{pmatrix} \mu_1 + \mu_2 & \mu_3 + \mu_4 \end{pmatrix}'$.

CHAPTER 6

MULTIVARIATE ANALYSIS OF VARIANCE

6.1 INTRODUCTION

This chapter deals with multivariate extensions of the procedures we discussed in Chapter 2 under Analysis of Variance. The extension of ANOVA to the multivariate case is commonly known as *Multivariate Analysis of Variance* and is abbreviated as MANOVA. We will also address the problem of extending the solution to the Behrens−Fisher problem to the case of comparing two multivariate normal populations. As will become clear later, some of the problems in the Analysis of Repeated Measures can also be handled in a MANOVA setting under alternative assumptions.

We encounter the MANOVA problem in comparing a number of populations when the underlying data consists of measurements on a number of dependent variables. The variables could be a number of distinct response variables observed for each of the experimental units or subjects. They could also be repeated measurements of one variable observed at distinct time points. In short, unlike in ANOVA, we are now dealing with a vector of possibly correlated responses or variables from each population as opposed to a single variable from each population. As in the previous chapter, the vector of observations may deal with the same characteristic, related characteristics, or completely different characteristics concerning the experimental unit. For example, the weight and height of one-year-old babies may be two characteristics we are dealing with. In a MANOVA problem with such data, for instance, one may be interested in comparing mean values of these variables for a number of ethnic groups in a certain income bracket. In this case, we would need a random sample of babies from each group from a certain population of subjects. In another application the only response variable of interest might be the blood pressure of some patients undergoing some treatments, and yet they are observed over time thus leading to a set of correlated variables. Yet in another application of different kind, one may wish to compare the crime rates of some cities of interest with data reported for a few months on a number of crimes.

6.2 COMPARING TWO MULTIVARIATE POPULATIONS

Suppose we have data from two multivariate populations, which we wish to compare. Let p be the dimension of the random vectors representing the two populations. The purpose of this section is to make inferences about the difference in the mean vectors of the two populations. Let $\mathbf{Y}_{11}, \ldots, \mathbf{Y}_{1m}$ be a sample of $p \times 1$ vector of responses from Population 1, and let $\mathbf{Y}_{21}, \ldots, \mathbf{Y}_{2n}$ be a $p \times 1$ vector of responses from Population 2. Assume that both populations are normally distributed and that the observations are independent; i.e.,

$$\begin{aligned} \mathbf{Y}_{1j} &\sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \quad j = 1, 2, \ldots, m \\ \mathbf{Y}_{2j} &\sim N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), \quad j = 1, 2, \ldots, n. \end{aligned} \tag{6.1}$$

In this section let us also make the additional assumption that the two covariance matrices are equal, an assumption that we will relax in the next section. Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ be the common covariance matrix. Denote by $\overline{\mathbf{Y}}_1$ and $\overline{\mathbf{Y}}_2$ the sample means of the two sets of data defined as in the previous section. The unbiased estimate of the covariance matrix based on all the data is

$$\mathbf{V} = \frac{\mathbf{W}}{m + n - 2},$$

where

$$\mathbf{W} = \sum_{j=1}^{m} (\mathbf{Y}_{1j} - \overline{\mathbf{Y}}_1)(\mathbf{Y}_{1j} - \overline{\mathbf{Y}}_1)' + \sum_{j=1}^{n} (\mathbf{Y}_{2j} - \overline{\mathbf{Y}}_2)(\mathbf{Y}_{2j} - \overline{\mathbf{Y}}_2)'.$$

Consider the problem of making inferences about the difference in the two mean vectors $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. This can be accomplished by taking an approach similar to that in Chapter 5 concerning the mean vector of a single population. It follows from (6.1) that the counterpart of the distribution given in (5.25) is

$$\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2 \sim N_p(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \left(\frac{1}{m} + \frac{1}{n}\right) \boldsymbol{\Sigma}),$$

whereas the counterpart of (5.26) is

$$\mathbf{W} \sim W_p(m+n-2, \boldsymbol{\Sigma}) \ . \tag{6.2}$$

Let $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ be the vector of mean differences. A statistic appropriate for testing hypotheses of the form

$$H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0 \tag{6.3}$$

and for constructing confidence regions for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ is the Hotelling's T^2 statistic

$$T^2 = \frac{mn}{m+n}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2 - \boldsymbol{\delta})'\mathbf{V}^{-1}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2 - \boldsymbol{\delta}), \tag{6.4}$$

whose distribution is given by

$$F = \frac{m+n-p-1}{p(m+n-2)}T^2 \sim F_{p,m+n-p-1} \ . \tag{6.5}$$

If the $\boldsymbol{\delta}$ is hypothesized as $\boldsymbol{\delta}_0$ in the computation of the Hotelling's T^2 statistic, then the above distribution would become a noncentral F distribution instead of the central F distribution. Now it is evident that the appropriate p-value for testing H_0 is

$$\begin{aligned} p &= \Pr(T^2 \geq \frac{mn}{m+n}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}_0)'\mathbf{v}^{-1}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}_0)) \\ &= 1 - H\left(r(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}_0)'\mathbf{v}^{-1}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}_0)\right), \end{aligned} \tag{6.6}$$

where

$$r = \frac{m+n-p-1}{p(m+n-2)}\frac{mn}{m+n},$$

$H = H_{p,m+n-p-1}$ is the cdf of the F distribution with p and $m+n-p-1$ degrees of freedom, $\mathbf{v}$ is the observed value of $\mathbf{V}$, and $\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2$ is the observed value of $\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2$. The $100\gamma\%$ confidence ellipsoid of $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ based the Hotelling's T^2 statistic is

$$r(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta})'\mathbf{v}^{-1}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}) \leq F_{p,m+n-p-1}(\gamma), \tag{6.7}$$

where $F_{p,m+n-p-1}(\gamma)$ is the γth quantile of the F distribution with p and $m+n-p-1$ degrees of freedom.

6.2.1 Simultaneous confidence intervals for differences in means

Scheffe-type simultaneous confidence intervals for individual components of $\boldsymbol{\delta}$ or any other linear function could be constructed as before by considering the class of all linear functions and by an argument as in (5.44). The counterpart of (5.45) obtained for linear combinations $\boldsymbol{\theta}_i = \mathbf{a}_i'\boldsymbol{\delta}$ is

$$\mathbf{a}_i'(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2) \pm k_\gamma \left((\frac{1}{m} + \frac{1}{n}) \mathbf{a}_i \mathbf{v} \mathbf{a}_i' \right)^{1/2}, \quad i = 1, 2, \ldots, r, \tag{6.8}$$

where

$$k_\gamma = \left(\frac{q(m+n-2)}{(m+n-q-1)} F_{q, m+n-q-1}(\gamma) \right)^{1/2}$$

and $q = p$ when the confidence level is ensured for any number of linear combinations. In constructing simultaneous confidence intervals for individual mean differences, the vectors $\mathbf{a}_i$ are defined to be vectors of zeros except for the ith element which is 1. It follows from the distributions

$$\mathbf{a}'(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2) \sim N_1(\boldsymbol{\theta}, (\frac{1}{m} + \frac{1}{n}) \mathbf{a} \boldsymbol{\Sigma} \mathbf{a}')$$

and

$$(m+n-2) \frac{\mathbf{a} \mathbf{V} \mathbf{a}'}{\mathbf{a} \boldsymbol{\Sigma} \mathbf{a}'} \sim \chi^2_{m+n-2} \tag{6.9}$$

that the form of the Bonferroni intervals could also be found by using (6.8) except for the definition of k_γ. It is now obvious that in the present situation k_γ should be computed as

$$k_\gamma = t_{m+n-2}(1 - \alpha/2r), \tag{6.10}$$

where $\alpha = 1 - \gamma$.

Example 6.1. Comparison of two diet plans

In a comparison of two diet plans, 11 men in a certain weight category were placed on diet Plan 1 and 13 other men in the same category were placed on diet Plan 2. Their weights were measured just before the diet and then later in one month and two months after being on the diets. Table 6.1 below shows the observed data from this repeated measures experiment, which one can analyze by alternative methods under alternative assumptions. Here let us analyze the data by applying the above methods.

Also shown in Table 6.1 are the sample means at three time points for each of the two diet plans. A plot of the sample means is shown in Figure 6.1. The figure indeed suggests that diet Plan 2 had a significant effect in reducing the weights. It is also clear from the figure that, even though all experimental subjects came from the same weight category, subjects placed under diet Plan 2 had substantially higher mean weight to start with. The question is whether or not the weight loss is statistically significant. This we can establish by applying the Hotelling T^2 test. In this application, however, it makes more sense to compare the weight losses rather than the weights themselves. This may also reduce the among-subject variation. Table 6.2 shows the weight loss of individual subjects in the two diet plans, obtained by subtracting weights after the diet from the weights before the diets.

Table 6.1 Subject weights (in pounds)before and after the diet

Diet Plan	Subject	Before	Month 1	Month 2
1	1	169	166	166
1	2	152	157	154
1	3	167	168	167
1	4	150	149	151
1	5	167	160	159
1	6	151	143	142
1	7	153	158	154
1	8	156	160	158
1	9	163	169	164
1	10	153	156	154
1	11	157	150	149
1	Means:	158.0	157.8	156.2
2	12	183	174	172
2	13	160	157	154
2	14	158	156	157
2	15	167	156	159
2	16	150	145	146
2	17	171	172	170
2	18	166	162	162
2	19	159	151	151
2	20	183	161	164
2	21	168	157	157
2	22	163	156	158
2	23	176	158	156
2	24	165	162	160
2	Means:	166.9	159.0	158.9

The mean vectors of weight losses for the two diet plans are $\overline{\mathbf{y}}_1 = (0.182, 1.819)$ and $\overline{\mathbf{y}}_2 = (7.846, 7.923)$, respectively. The sample covariances computed from the two data sets are

$$V_1 = \begin{pmatrix} 28.36 & 22.04 \\ 22.04 & 19.36 \end{pmatrix} \text{ and } V_2 = \begin{pmatrix} 42.31 & 37.24 \\ 37.24 & 36.24 \end{pmatrix},$$

respectively. The pooled sample covariance matrix computed from them, under the assumption of equal covariance matrices, is

$$V = \begin{pmatrix} 35.97 & 30.33 \\ 30.33 & 28.57 \end{pmatrix}.$$

Notice that, even after subtracting the weight of each subject before the experiment, the two data vectors are still correlated. The p-value for testing the

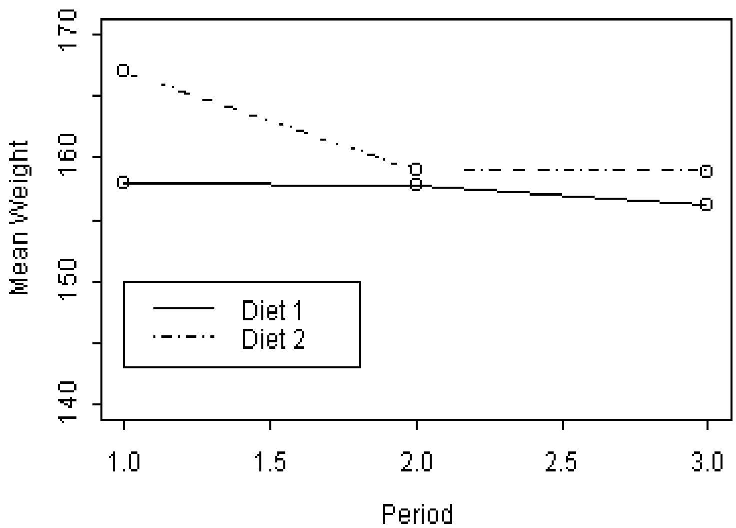

Figure 6.1 Mean weights by diet plan

Table 6.2 Weight loss (in pounds)

Diet Plan 1		Diet Plan 2	
Month 1	Month 2	Month 1	Month 2
3	3	9	11
-5	-2	3	6
-1	0	2	1
1	-1	11	8
7	8	5	4
8	9	-1	1
-5	-1	4	4
-4	-2	8	8
-6	-1	22	19
-3	-1	11	11
7	8	7	5
		18	20
		3	5

equality of the two mean vectors can now be computed by applying formula (6.6). The p-value of 0.020 computed in this manner suggests the rejection of the null hypothesis. Hence, the expected superiority of Diet Plan 2 is indeed statistically significant. The 95% simultaneous confidence intervals for the difference in individual means, namely (-13.6, -1.734) for the difference in means after one month, and (-11.4, -0.82) for the difference in means after two months, computed with the Bonferroni level adjustment, further support the conclusion.

6.3 MULTIVARIATE BEHRENS−FISHER PROBLEM

In the treatment of the previous section, we assumed that the covariance matrices of the two populations being compared are equal. The problem of comparing multivariate normal mean vectors without that assumption is commonly known as the Multivariate Behrens−Fisher Problem. There are a number of approximate and exact solutions to the problem in the literature on this subject. A number of studies, including the recent article by Christensen and Rencher (1997), showed that many approximate solutions tend to have Type I error well exceeding the nominal level. Bennett (1963) gave an exact solution to the problem of testing the equality of two normal mean vectors by taking the approach of Scheffe (1943) in the case of univariate normal populations. It is well known that this solution is far from being most powerful, because it is not based on sufficient statistics. Johnson and Weerahandi (1988) provided a Bayesian solution to the problem that does not suffer from that drawback. By taking the generalized p-value approach, Gamage (1997) provided an upper bound for the p-value, and Gamage, Mathew, and Weerahandi (2004) obtained exact generalized p-values and confidence regions without using Bayesian arguments. The confidence region given by the generalized p-value approach is also numerically equivalent to the Bayesian solution under the noninformative prior.

To outline their results, consider the problem of testing hypotheses of the form $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$, where $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. A set of sufficient statistics for the problem are

$$\overline{\mathbf{Y}}_1, \overline{\mathbf{Y}}_2, \mathbf{W}_1 = \sum_{j=1}^{m} (\mathbf{Y}_{1i} - \overline{\mathbf{Y}}_1)(\mathbf{Y}_{1i} - \overline{\mathbf{Y}}_1)', \tag{6.11}$$

and

$$\mathbf{W}_2 = \sum_{j=1}^{n} (\mathbf{Y}_{2j} - \overline{\mathbf{Y}}_2)(\mathbf{Y}_{2j} - \overline{\mathbf{Y}}_2)'.$$

Theses random vectors are independently distributed as

$$\overline{\mathbf{Y}}_i \sim N(\boldsymbol{\mu}_i, \frac{\boldsymbol{\Sigma}_i}{n_i}),\ \mathbf{W}_i \sim W_p(n_i - 1, \Sigma_i),\ i = 1,\ 2, \tag{6.12}$$

where $n_1 = m$ and $n_2 = n$. The unbiased estimates of the covariance matrices based on $\mathbf{W}_i$ are $\mathbf{S}_i = \mathbf{W}_i/(n_i - 1),\ i = 1,\ 2$. A class of solutions, non-Bayesian as well as Bayesian, could be derived based on observable random vectors in (6.12). Here only a natural solution, having relationships with the solutions to the univariate problem and the solution under equal covariance matrices, is presented. The generalized p-value that Gamage, Mathew, and Weerahandi (2004) obtained for testing hypothesis (6.3) is

$$p = \Pr(T(\mathbf{v}) \geq (\overline{\mathbf{y}}_{\mathbf{1}} - \overline{\mathbf{y}}_{\mathbf{2}} - \boldsymbol{\delta}_{\mathbf{0}})' \left(\frac{\mathbf{s}_1}{n_1} + \frac{\mathbf{s}_2}{n_2} \right)^{-1} (\overline{\mathbf{y}}_{\mathbf{1}} - \overline{\mathbf{y}}_{\mathbf{2}} - \boldsymbol{\delta}_{\mathbf{0}})), \qquad (6.13)$$

where

$$T(\mathbf{v}) = \mathbf{Z}'(\mathbf{v}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{v}_1^{1/2} + \mathbf{v}_2^{1/2}\mathbf{R}_2^{-1}\mathbf{v}_2^{1/2})\mathbf{Z}$$

is a test variable, $\mathbf{v}_i$ is the observed value of

$$\mathbf{V}_i = (n_i - 1) \left(\frac{\mathbf{s}_1}{n_1} + \frac{\mathbf{s}_2}{n_2} \right)^{-1/2} \frac{\mathbf{S}_i}{n_i} \left(\frac{\mathbf{s}_1}{n_1} + \frac{\mathbf{s}_2}{n_2} \right)^{-1/2}, \qquad i = 1,\ 2, \qquad (6.14)$$

and the probability is computed in terms of

$$\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_p) \quad \text{and} \quad \mathbf{R}_i \sim W_p(n_i - 1, \mathbf{I}_p)\ , \quad i = 1,\ 2. \qquad (6.15)$$

6.3.1 Derivation of the Gamage—Mathew—Weerahandi test

Define a set of transformed covariance matrices as

$$\Lambda_i = g(\mathbf{s})^{-1/2} \frac{\boldsymbol{\Sigma}_i}{n_i} g(\mathbf{s})^{-1/2}, \qquad i = 1,\ 2, \qquad (6.16)$$

where

$$g(\mathbf{s}) = \frac{\mathbf{s}_1}{n_1} + \frac{\mathbf{s}_2}{n_2} \qquad (6.17)$$

and by square root notation of a positive definite matrix we mean its square root constructed using the positive square roots of its eigenvalues. More precisely, if A is a positive definite covariance with the spectral decomposition

$$A = Q'DQ,$$

then its square root matrix is defined as

$$A^{1/2} = Q'D^{1/2}Q,$$

where Q is the matrix formed by eigenvectors of A and D is the diagonal matrix with eigenvalues on its diagonal. To show that $T(\mathbf{v})$ is a test variable appropriate for testing the hypothesis (6.3), define the $\mathbf{Z}$ random vector as

$$\mathbf{Z} = (\Lambda_1 + \Lambda_2)^{-1/2} g(\mathbf{s})^{-1/2} (\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2 - \boldsymbol{\delta}_{\mathbf{0}}). \qquad (6.18)$$

It follows from (6.12) that, under H_0,

$$g(\mathbf{s})^{-1/2} (\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2 - \boldsymbol{\delta}_{\mathbf{0}}) \sim N(\mathbf{0}, \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2)$$

and hence $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_p)$ as required in (6.15). From (6.12) we also get

$$\mathbf{V}_i = (n_i - 1)g(\mathbf{s})^{-1/2}\frac{\mathbf{S}_i}{n_i}g(\mathbf{s})^{-1/2} \sim W_p(n_i - 1, \mathbf{\Lambda}_i), \tag{6.19}$$

and in turn $\mathbf{v}_i^{-1/2}\mathbf{V}_i\mathbf{v}_i^{-1/2} \sim W_p(n_i - 1, \mathbf{v}_i^{-1/2}\mathbf{\Lambda}_i\mathbf{v}_i^{-1/2})$. Hence we can define the $\mathbf{R}_i$ matrix appearing in (6.15) as

$$\begin{aligned}\mathbf{R}_i &= (\mathbf{v}_i^{-1/2}\mathbf{\Lambda}_i\mathbf{v}_i^{-1/2})^{-1/2}(\mathbf{v}_i^{-1/2}\mathbf{V}_i\mathbf{v}_i^{-1/2})(\mathbf{v}_i^{-1/2}\mathbf{\Lambda}_i\mathbf{v}_i^{-1/2})^{-1/2} &(6.20)\\ &\sim W_p(n_i - 1, \mathbf{I}_p), \quad i = 1,\ 2. &(6.21)\end{aligned}$$

Notice that at the observed point of the sample space, $\mathbf{R}_i$ takes on the value $\mathbf{v}_i^{-1/2}\Lambda_i\mathbf{v}_i^{-1/2}$. Consequently, when $\mathbf{Z}$, $\mathbf{R}_1$, and $\mathbf{R}_2$ are defined as in (6.17) and (6.20), the observed value of the potential test variable

$$T(\mathbf{v}) = \mathbf{Z}'(\mathbf{v}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{v}_1^{1/2} + \mathbf{v}_2^{1/2}\mathbf{R}_2^{-1}\mathbf{v}_2^{1/2})\mathbf{Z} \tag{6.22}$$

becomes

$$\begin{aligned}t &= (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}_0)'g(\mathbf{s})^{-1/2}\mathbf{\Lambda}_{sum}^{-1/2}\mathbf{\Lambda}_{sum}\mathbf{\Lambda}_{sum}^{-1/2}g(\mathbf{s})^{-1/2}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}_0)\\ &= (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}_0)'g(\mathbf{s})^{-1}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}_0)\ , &(6.23)\end{aligned}$$

where $\Lambda_{sum} = \Lambda_1 + \Lambda_2$. Moreover, the distribution $T(\mathbf{v})$ is free of unknown parameters and $\Pr(T(\mathbf{v}) \geq t)$ increases for any deviation from the null hypothesis, given $\mathbf{R}_1$ and $\mathbf{R}_2$. Also note that $T(\mathbf{v})$ is a positive definite quadratic form leading to a distribution having linear combinations of noncentral chi-squared random variates with the noncentrality parameter becoming zero only under H_0. This fact will become further clear later in this section. Hence, $T(\mathbf{v})$ is a test variable appropriate for testing H_0 and its generalized p-value is given by (6.13). Gamage, Mathew, and Weerahandi (2004) also showed that this test is invariant under transformations, which leaves the testing problem invariant.

6.3.2 Computing the generalized p-value

In computing the generalized p-value given by (6.13) it is convenient and numerically much more advantageous if it is expressed in terms of univariate random variables. To do this, define

$$\mathbf{U}_i = \mathbf{v}_i^{1/2}\mathbf{R}_i^{-1}\mathbf{v}_i^{1/2} \sim W_p(n_i - 1, \mathbf{v}_i^{-1}), \quad i = 1,\ 2 \tag{6.24}$$

from (6.20). In terms of these random variables, the test variable can be expressed as

$$T(\mathbf{v}) = \mathbf{Z}'(\mathbf{U}_1^{-1} + \mathbf{U}_2^{-1})\mathbf{Z}. \tag{6.25}$$

Now define

$$Q_i = \frac{\mathbf{Z}'\mathbf{v}_i\mathbf{Z}}{\mathbf{Z}'\mathbf{U}_i^{-1}\mathbf{Z}} \sim \chi^2_{n_i - p}\ , \quad i = 1,\ 2. \tag{6.26}$$

That these are chi-square random variables follow from the fact that, conditionally given $\mathbf{Z}$, $Q_i \sim \chi^2_{n_i-p}$, which does not depend on $\mathbf{Z}$, and hence it is also the unconditional distribution. Consequently,

$$T(\mathbf{v}) = \mathbf{Z}'(\mathbf{U}_1^{-1}+\mathbf{U}_2^{-1})\mathbf{Z} = \frac{1}{Q_1}\mathbf{Z}'\mathbf{v}_1\mathbf{Z} + \frac{1}{Q_2}\mathbf{Z}'\mathbf{v}_2\mathbf{Z}, \qquad (6.27)$$

where $\mathbf{Z} \sim N(0, I_p)$, Q_1 and Q_2 are independent. This representation of the test variable enables us to compute the p-value by Monte Carlo integration with random numbers generated from independent normal and chi-squared distributions.

Johnson and Weerahandi (1988) discussed various methods of computing the probabilities associated with this type of random variables to be exact to any desired level of accuracy. Therefore, the procedures mentioned by Johnson and Weerahandi (1988) could be used for computing the generalized p-value also. However, the expression given by Johnson and Weerahandi (1988) involves an infinite series, although it reduces to a finite summation when limited accuracy is required, as always the case in numerical computations. Gamage, Mathew, and Weerahandi (2004) developed an expression to compute the generalized p-value involving the distributions of only a finite number of univariate random variables. To present their expression, first note that

$$\frac{\mathbf{v}_1}{n_1-1} + \frac{\mathbf{v}_2}{n_2-1} = I_p.$$

Hence,

$$\mathbf{v}_2 = (n_2-1)(\mathbf{I}_p - \frac{\mathbf{v}_1}{n_1-1})$$

and T defined by (6.27) can be expressed as

$$T = \frac{1}{Q_1}\mathbf{Z}'\mathbf{v}_1\mathbf{Z} + \frac{n_2-1}{Q_2}\mathbf{Z}'(\mathbf{I}_p - \frac{\mathbf{v}_1}{n_1-1})\mathbf{Z}. \qquad (6.28)$$

This representation of T allows us to use the diagonalization matrices of $\mathbf{v}_1$ to diagonalize $\mathbf{v}_2$ as well. Hence the test variable can be expressed as

$$T(\mathbf{v}) = \frac{1}{Q_1}\sum_{i=1}^{p}\lambda_i Z_i^2 + \frac{n_2-1}{Q_2}\sum_{i=1}^{p}(1 - \frac{\lambda_i}{n_1-1})Z_i^2, \qquad (6.29)$$

where $\lambda_i,\ i = 1, 2, ..., p$. In (6.29), each Z_i^2, $i = 1, 2, ..., p$, have a chi-squared distributions with 1 degree of freedom, under H_0, and Q_1, Q_2, and the Z_i^2's are all independently distributed. Thus, the representation (6.29) could be used in computing the generalized p-value in (6.13) by exact integration or by simulation.

Example 6.2. Comparison of two diet plans (continued)

Consider again the data set in Table 6.1. In Example 6.1 we compared the two diet plans under the assumption that the two covariance matrices are equal.

The sample covariance matrices given by the transformed data in Table 6.2 was

$$V_1 = \begin{pmatrix} 28.36 & 22.04 \\ 22.04 & 19.36 \end{pmatrix} \quad \text{and} \quad V_2 = \begin{pmatrix} 42.31 & 37.24 \\ 37.24 & 36.24 \end{pmatrix}.$$

According to the above sample covariances the assumption made by the Hotelling's T^2 test is not very reasonable. In any case, with the above results we can carry out the test without relying on an additional assumption. The generalized p-value given by (6.13) is $p = 0.031$. Hence, we come to the same conclusion as before that the mean reduction in weights due to the two diet plans are significantly different.

6.3.3 Generalized Confidence Regions

Generalized confidence regions for $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ is easily deduced from the formula (6.13) on the generalized p-value or could be obtained from the same $T(\mathbf{v})$ on which p-value was based on, when the $\mathbf{Z}$ random vector is refined as

$$\mathbf{Z} = (\Lambda_1 + \Lambda_2)^{-1/2} g(\mathbf{s})^{-1/2} (\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2 - \boldsymbol{\delta}) \sim N(\mathbf{0}, \mathbf{I}_p). \tag{6.30}$$

It is easily seen that $T(\mathbf{v}) = \mathbf{Z}'(\mathbf{v}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{v}_1^{1/2} + \mathbf{v}_2^{1/2}\mathbf{R}_2^{-1}\mathbf{v}_2^{1/2})\mathbf{Z}$ is a generalized pivotal quantity that is free of unknown parameters and that its observed value is

$$t = (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta})' g(\mathbf{s})^{-1} (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}). \tag{6.31}$$

Now it is evident that the $100\gamma\%$ confidence ellipsoid of $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ is given by

$$(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta})' \mathbf{g}(\mathbf{s})^{-\mathbf{1}} (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta}) \le F_T(\gamma), \tag{6.32}$$

where $F_T(\gamma)$ is the γth percentile of the distribution of $T(\mathbf{v})$. Unlike other solutions in the literature, on one hand the generalized confidence region given by (6.32) has the advantage that it is numerically equivalent to the Bayesian confidence region given by Johnson and Weerahandi (1988), under the usual noninformative prior. On the other hand unlike the Bayesian solution, the generalized p-value easily extends to the case of more than two multivariate normal populations, a task that we will undertake in the following section.

6.3.4 Simultaneous Confidence Intervals

Simultaneous confidence intervals or tests for individual components of $\boldsymbol{\delta}$ or any other linear function can be deduced from the fact that

$$\mathbf{a}'(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2 - \boldsymbol{\delta}) \sim N(0, \mathbf{a}'(\frac{\boldsymbol{\Sigma}_1}{n_1} + \frac{\boldsymbol{\Sigma}_2}{n_2})\mathbf{a}). \tag{6.33}$$

In this case we can define a test variable as

$$T = \mathbf{a}'(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2 - \boldsymbol{\delta}) \left(\frac{\mathbf{a}'(\frac{\mathbf{v}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{v}_1^{1/2}}{n_1} + \frac{\mathbf{v}_2^{1/2}\mathbf{R}_2^{-1}\mathbf{v}_2^{1/2}}{n_2})\mathbf{a}}{\mathbf{a}'(\frac{\boldsymbol{\Sigma}_1}{n_1} + \frac{\boldsymbol{\Sigma}_2}{n_2})\mathbf{a}} \right)^{1/2} \quad (6.34)$$

$$= Z\sqrt{\mathbf{a}'(\frac{\mathbf{v}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{v}_1^{1/2}}{n_1} + \frac{\mathbf{v}_2^{1/2}\mathbf{R}_2^{-1}\mathbf{v}_2^{1/2}}{n_2})\mathbf{a}} \quad (6.35)$$

$$= Z\sqrt{\mathbf{a}'(\frac{\mathbf{v}_1}{n_1 Q_1} + \frac{\mathbf{v}_2}{n_2 Q_2})\mathbf{a}}, \quad (6.36)$$

where $Z \sim N(0,1)$, $Q_i \sim \chi^2_{n_i - p}$, and, $\mathbf{R}_i$ is redefined as

$$\begin{aligned} \mathbf{R}_i &= (\mathbf{v}_i^{-1/2}\boldsymbol{\Sigma}_i\mathbf{v}_i^{-1/2})^{-1/2}(\mathbf{v}_i^{-1/2}\mathbf{V}_i\mathbf{v}_i^{-1/2})(\mathbf{v}_i^{-1/2}\boldsymbol{\Sigma}_i\mathbf{v}_i^{-1/2})^{-1/2} \\ &\sim W_p(n_i - 1, \mathbf{I}_p), \quad i = 1,\ 2. \end{aligned}$$

From the representation (6.35) it is evident that the distribution T is free of unknown parameters, and from (6.34) it is clear that the observed value of T is $\mathbf{a}'(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2 - \boldsymbol{\delta})$. Hence, the two-sided $100\gamma\%$ confidence intervals of $\theta = \mathbf{a}'\boldsymbol{\delta}$ are of the form

$$\mathbf{a}'(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2) \pm k_\gamma(\mathbf{v}), \quad (6.37)$$

where $k_\gamma(\mathbf{v})$ is the $[(1+\gamma)/2]$th quantile of the distribution of T. If simultaneous intervals are to be carried out, then formula (6.37) should be applied with the Bonferroni adjustment to $\alpha = 1 - \gamma$, as we did in the previous section.

6.4 MANOVA WITH EQUAL COVARIANCES

Next consider the case where we have data from a number of multivariate populations of dimension p. Let $\mathbf{Y}_{i1}, \cdots, \mathbf{Y}_{in_i}$ be a sample of $p \times 1$ vector of responses from Population i; $i = 1, \ldots, I$. Continuing with the normal theory we assume that

$$\mathbf{Y}_{ij} \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}); \quad i = 1, \ldots, I;\ j = 1, \ldots, n_i. \quad (6.38)$$

Table 6.3 presents an example of a typical data set as discussed by Seber (1984). The last raw of the table shows the sample means of each column of data.

As in the previous section, the obvious unbiased estimate of the mean vector $\boldsymbol{\mu}_i$ is the sample mean vector $\overline{\mathbf{Y}}_i = \sum \mathbf{Y}_{ij}/n_i$. These means are presented in the last raw of Table 6.3. As will become clear later, the unbiased estimate of the common covariance matrix $\boldsymbol{\Sigma}$ is the pooled sample covariance matrix

$$\mathbf{V} = \frac{\mathbf{E}}{N - I}, \quad \text{where } \mathbf{E} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(\mathbf{Y}_{ij} - \overline{\mathbf{Y}}_i)(\mathbf{Y}_{ij} - \overline{\mathbf{Y}}_i)' \quad (6.39)$$

Table 6.3 Logarithm of measurements on anteater skulls at 3 localities

Minas Graes, Brazil			Matto Grosso, Brazil			Sanra Cruz, Bolivia		
$\mathbf{y}_1$	$\mathbf{y}_2$	$\mathbf{y}_3$	$\mathbf{y}_1$	$\mathbf{y}_2$	$\mathbf{y}_3$	$\mathbf{y}_1$	$\mathbf{y}_2$	$\mathbf{y}_3$
2.068	2.070	1.580	2.045	2.054	1.580	2.093	2.098	1.653
2.068	2.074	1.602	2.076	2.088	1.602	2.100	2.106	1.623
2.090	2.090	1.613	2.090	2.093	1.643	2.104	2.101	1.653
2.097	2.093	1.613	2.111	2.114	1.643			
2.117	2.125	1.663						
2.140	2.146	1.681						
2.097	2.100	1.625	2.080	2.87	1.617	2.099	2.102	1.643

is the within-population sum of cross products and $N = \sum n_i$. For the data in Table 6.3 concerning 3 populations, the pooled sample covariance matrix is

$$\mathbf{V} = \begin{pmatrix} 0.000634 & 0.000624 & 0.000762 \\ 0.000624 & 0.000635 & 0.000761 \\ 0.000762 & 0.000761 & 0.001097 \end{pmatrix}.$$

An important and basic problem of interest in many applications involving a number multivariate populations is whether the population means are the same or not; i.e., testing of the null hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \cdots = \boldsymbol{\mu}_I \ . \tag{6.40}$$

It is important that one first compares the equality of all mean vectors even if the ultimate goal is to identify a certain population with most desirable characteristics that the mean vector represent. If the hypothesis is rejected, we can then proceed to perform pairwise comparisons as well as inferences concerning individual populations with less concern about the Type I error.

The model can also be written as

$$\mathbf{Y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ij}, \quad \text{where } \boldsymbol{\epsilon}_{ij} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}). \tag{6.41}$$

Define $n_i \times p$ matrix of data from ith population as

$$\mathbf{Y}_i = \begin{pmatrix} \mathbf{Y}'_{i1} \\ \mathbf{Y}'_{i2} \\ \vdots \\ \mathbf{Y}'_{in_i} \end{pmatrix}.$$

Then, piling all dependent variables into a single $N \times p$ matrix $\mathbf{Y}$, the model can be expressed as a multivariate regression model as

$$\mathbf{Y} = \mathbf{XB} + \varepsilon, \tag{6.42}$$

where

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}'_I \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \dots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \dots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{n_I} & \mathbf{0}_{n_I} & \dots & \mathbf{1}_{nI} \end{pmatrix}, \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} \boldsymbol{\mu}'_1 \\ \boldsymbol{\mu}'_2 \\ \vdots \\ \boldsymbol{\mu}'_I \end{pmatrix}$$

is an $I \times p$ matrix of parameters formed by the mean vectors of interest. The MLE of $\mathbf{B}$ with a general design matrix $\mathbf{X}$ is $\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and it reduces to a matrix of sample means in the above case. It can also be shown that the MLE of $\boldsymbol{\Sigma}$ is $\widehat{\boldsymbol{\Sigma}} = \mathbf{Q}/\mathbf{N}$, where $\mathbf{Q} = \mathbf{Y}'(\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')^{-1}\mathbf{Y}$, which reduces to $\mathbf{E}$ under the model (6.38). For details and derivations of these and the distributional results given below, the reader is referred to Muirhead (1982), Seber (1984), and Anderson (1984).

Notice that the hypothesis of equal population means can be expressed in the form

$$H_0 : \mathbf{CB} = \mathbf{0}, \tag{6.43}$$

where

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix} \tag{6.44}$$

is an $(I-1) \times I$ matrix. In fact, we can use appropriately defined $\mathbf{C}$ matrix to test various hypotheses involving all or some of the mean vectors.

6.4.1 Multiple comparisons

The above representation is especially useful in follow up multiple comparisons after testing the equality of all mean vectors. With that representation, the distributional results follow from those available for general linear models. In fact, hypotheses involving even individual elements of mean vectors could be handled by considering more general hypotheses involving double linear combinations of the form

$$H_0 : \mathbf{CBD} = \mathbf{0}, \tag{6.45}$$

and a general design matrix $\mathbf{X}$ as well, where $\mathbf{C}$ is a $c \times I$ matrix of rank $c \leq I$, and $\mathbf{D}$ is a $p \times d$ matrix of rank $d \leq p$, formed by a number of comparisons of interest. For example, to test the hypothesis that the first coefficient of each of the I populations is identical, we choose the $\mathbf{C}$ matrix as in (6.44) and define $\mathbf{D} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}'$ as a $p \times 1$ vector. To compare only the first two populations we set, $\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \end{pmatrix}$.

In the general case, the null hypothesis in (6.45) can be tested based on the distributions given by

$$\mathbf{E} = \mathbf{D}'\mathbf{Q}\mathbf{D} \sim W_d(e, \mathbf{D}'\mathbf{\Sigma}\mathbf{D}), \tag{6.46}$$
$$\mathbf{H} = (\mathbf{C}\widehat{\mathbf{B}}\mathbf{D})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\widehat{\mathbf{B}}\mathbf{D}) \sim W_d(h, \mathbf{D}'\mathbf{\Sigma}\mathbf{D}), \tag{6.47}$$

which are valid under H_0, where $e = N - I$, $h = c$. If the null hypothesis is not true, the random matrix $\mathbf{H}$ has a noncentral Wishart distribution with noncentrality parameter matrix $\Psi = (\mathbf{CBD})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{CBD})$. These results and (6.45) allow us to test not only the hypotheses concerning mean vectors but also those involving components of individual mean vectors. In testing the particular hypothesis (6.40) of equality of mean vectors, the matrices $\mathbf{E}$ and $\mathbf{H}$ reduce to the multivariate extension of the within population and between population sums of cross products given by the ANOVA table shown below.

Table 6.4 Multivariate ANOVA

Source	Matrix	DF
Between populations	$\mathbf{H} = \sum_{i=1}^{I} n_i(\overline{\mathbf{Y}}_i - \overline{\mathbf{Y}})(\overline{\mathbf{Y}}_i - \overline{\mathbf{Y}})'$	$I - 1$
Within populations (error)	$\mathbf{E} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(\mathbf{Y}_{ij} - \overline{\mathbf{Y}}_i)(\mathbf{Y}_{ij} - \overline{\mathbf{Y}}_i)'$	$N - I$

In testing the equality of mean vectors of the populations, the distributions of $\mathbf{E}$ and $\mathbf{H}$ also reduce to

$$\mathbf{E} \sim W_p(N - I, \mathbf{\Sigma}), \tag{6.48}$$
$$\mathbf{H} \sim W_p(I - 1, \mathbf{\Sigma}). \tag{6.49}$$

The literature [cf. Seber (1984)] on multivariate analysis provides multiple procedures for testing (6.45) based on the ordered eigenvalues of the matrix $\mathbf{HE}^{-1}$, say $e_1, e_1, \ldots, e_h$. This is because, unlike the two population problem, there is no uniformly most powerful test in the multivariate case. Some widely used tests are the Wilks likelihood ratio test with the test statistic

$$T_W = \prod_{k=1}^{h} \frac{1}{1 + e_k} = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} = \frac{1}{|\mathbf{I}_d + \mathbf{HE}^{-1}|}, \tag{6.50}$$

Roy's largest root test $T_R = e_1$, Lawley–Hotelling's test $T_{LH} = \sum e_k$, and the Bartlett–Nanda–Pillai test $T_{BNP} = \sum e_k/(1 + e_k)$. There is no clear winner among these tests in terms of the power of tests. The p-value of any of these tests can be computed by simulating the distributions using standard normal random variates. As far as the exact distributions of the test statistics are concerned, only the likelihood ratio test is tractable for any number of populations being compared. In general, the likelihood ratio test has a U distribution with d, h, and e degrees of freedom,

$$T_W \;= U \sim U_{d,h,e}.$$

The p-value for testing the null hypotheses of the form (6.45) in general and the hypothesis (6.40) of particular interest can be computed as

$$p = \Pr(T_W \leq t_W) = F_U(t_W), \tag{6.51}$$

where F_U is the cdf of the U distribution with d, h and e degrees of freedom, and t_W is the observed value of the Wilks statistic. Asymptotically the statistic, $-(e - (d - h + 1)/2)\,\log(T_W)$ has a chi-squared distribution with dh degrees of freedom. Although this approximation is very good in computing some critical values, in this computer age there is no real need to resort to such asymptotic results, especially in computing p-values regardless of the observed value of the statistic. The probabilities of the U distribution can be evaluated by means of alternative methods given in the last section of Chapter 5. In particular, when the d or h parameter is less than 3, its distribution can be transformed into an F distribution. When d is large, perhaps the best way to compute the p-value is to generate a large number of random numbers from each of the independent beta distributions and computing the fraction of times that the inequality appearing in (6.51) is satisfied.

To illustrate the importance of the above results, suppose, for instance, we are interested in comparing only the first two populations. In this case, $\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \end{pmatrix}$ and $\mathbf{D} = \mathbf{I}_p$ so that $c = 1$ and $d = p$. Moreover, $h = c = 1$ and $\mathbf{H} = (1/n_1 + 1/n_2)^{-1}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)'$ is of rank 1. Hence, $U = 1/(1+\mathrm{tr}(\mathbf{HE}^{-1})$ and in turn

$$\begin{aligned} F &= \frac{1-U}{U}\frac{e+1-p}{p} \\ &= \mathrm{tr}(\mathbf{HE}^{-1})\frac{N-p-I+1}{p} \\ &\sim F_{p,N-p-I+1}. \end{aligned} \tag{6.52}$$

Using the properties of the trace operator, we get

$$\mathrm{tr}(\mathbf{HE}^{-1}) = \frac{n_1 n_2}{n_1 + n_2}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2)'\mathbf{E}^{-1}(\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_2). \tag{6.53}$$

Now it is evident that the F-statistic given by (6.52) is the same as the one used in the two population problem, (6.5) except that now we are using data from all I populations to estimate the covariance matrix. As a result, the second degrees of freedom of the F distribution is now $N-p-I+1$ instead of n_1+n_2-p-1. Furthermore, the two F-statistics and the degrees of freedom are identical when $I=2$.

Example 6.3. Comparison of anteater skulls (continued)

Consider the data in Table 6.3 concerning three populations. The observed matrices of between populations and within population sums of cross products are

$$\mathbf{H} = \begin{pmatrix} 0.000806 & 0.000623 & 0.000750 \\ 0.000623 & 0.000482 & 0.000586 \\ 0.000750 & 0.000586 & 0.001184 \end{pmatrix}$$

and

$$\mathbf{E} = \begin{pmatrix} 0.006342 & 0.006241 & 0.007615 \\ 0.006241 & 0.006352 & 0.007613 \\ 0.007615 & 0.007613 & 0.010967 \end{pmatrix}$$

with degrees of freedom $h=2$ and $e=10$, respectively. The likelihood ratio computed using is $T_W = U = |\mathbf{E}|/|\mathbf{E}+\mathbf{H}| = 0.601437$. In comparing the three populations we can transform the U-statistic into the F-statistic

$$F = \frac{1-U^{1/2}}{U^{1/2}}\frac{8}{3} \sim F_{6,16}.$$

The observed value of the F-statistic is 0.7719. Hence, the p-value for testing the equality of three population mean vectors is 0.6032. Therefore, we can conclude that there is no significant difference in the three population mean vectors.

6.4.2 Simultaneous confidence regions

A particular case of (6.47) is that

$$\begin{aligned} \mathbf{E} &= \mathbf{N}\widehat{\mathbf{\Sigma}} \sim W_p(N-I, \mathbf{\Sigma}) \\ \mathbf{H} &= (\widehat{\mathbf{B}}-\mathbf{B})'(\mathbf{X}'\mathbf{X})(\widehat{\mathbf{B}}-\mathbf{B}) \sim W_p(I, \mathbf{\Sigma}), \end{aligned}$$

where $\widehat{\mathbf{\Sigma}} = \mathbf{Y}'(\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')^{-1}\mathbf{Y}/\mathbf{N}$ is the MLE of $\mathbf{\Sigma}$. Confidence regions for the matrix of all means, $\mathbf{B}$, follow immediately from this result. As we discussed in Chapter 5, the joint $100\gamma\%$ confidence region for $\mathbf{B}$ can be found as

$$\{\mathbf{B} \mid |\mathbf{I}_p + \frac{1}{N}(\widehat{\mathbf{B}}-\mathbf{B})'(\mathbf{X}'\mathbf{X})(\widehat{\mathbf{B}}-\mathbf{B})\widehat{\mathbf{\Sigma}}^{-1}| \le \kappa_\gamma\}, \qquad (6.54)$$

where κ_γ is the $(1-\gamma)$th quantile of the U distribution with p, I, and $N-I$ degrees of freedom. $100\gamma\%$ confidence regions for any desired contrast matrix

$$\mathbf{\Phi} = \mathbf{CBD}$$

of dimension $c \times d$ can be similarly constructed using the general form of (6.47) as

$$\{\mathbf{\Phi} \mid |\mathbf{I}_p + (\widehat{\mathbf{\Phi}} - \mathbf{\Phi})'((\mathbf{C}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{C}')^{-\mathbf{1}})(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})\mathbf{E}^{-1}| \leq \kappa_\gamma\}, \tag{6.55}$$

where $\mathbf{E} = \mathbf{D}'\mathbf{Y}'(\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')^{-1}\mathbf{Y}\mathbf{D}$, and κ_γ is the $(1-\gamma)$th quantile of the U distribution with d, h, and e degrees of freedom.

Of particular interest is the problem of constructing simultaneous confidence intervals of linear combinations of means. If comparisons of a certain component of each of the populations had been planned, then we set $d = 1$ in (6.55) so that it reduces to a Scheffe-type confidence interval based on the F-statistic. If there were r pre-planned linear combination of interest of the form $\theta_i = \mathbf{a}_i'\mathbf{\Phi}\mathbf{b}_i$, then Bonferroni-type intervals can be constructed using the distributional results

$$\widehat{\theta}_i = \mathbf{a}_i'\widehat{\mathbf{\Phi}}\mathbf{b}_i \sim \mathbf{N}(\theta_i, (\mathbf{a}_i'\mathbf{C}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{C}'\mathbf{a_i})(\mathbf{b}_i'\mathbf{D}'\mathbf{\Sigma}\mathbf{D}\mathbf{b}_i))$$

and

$$\frac{\mathbf{b}_i'\mathbf{E}\mathbf{b}_i}{\mathbf{b}_i'\mathbf{D}'\mathbf{\Sigma}\mathbf{D}\mathbf{b}_i} \sim \chi^2_{N-I},$$

which lead to a t-statistic. The $100\gamma\%$ simultaneous confidence intervals constructed in this manner for θ_i, $i = 1, \ldots, r$ are

$$\widehat{\theta}_i \pm t_\nu(1 - \frac{\alpha}{2r})\left(\frac{v_{ab}}{e}\right)^{1/2}, \; i = 1, 2, \ldots, r, \tag{6.56}$$

where $v_{ab} = (\mathbf{a}_i'\mathbf{C}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{C}'\mathbf{a_i})(\mathbf{b}_i'\mathbf{E}\mathbf{b}_i)$ and $\alpha = 1 - \gamma$. In constructing confidence intervals for linear combinations directly based on the matrix of means $\mathbf{B}$ as defined in (6.42), $\widehat{\theta}_i = \mathbf{a}_i'\widehat{\mathbf{B}}\mathbf{b}_i$ and $v_{ab} = (\sum a_{ij}^2/n_j)(\mathbf{b}_i'\mathbf{E}\mathbf{b}_i)$ with $\mathbf{E}$ appearing in the MANOVA table.

Based on the maximum root criterion, Roy and Bose (1966) derived simultaneous confidence intervals to be valid for any number of linear combinations. The $100\gamma\%$ confidence bounds for $\theta_{\mathbf{ab}}$ given by the maximum root method is

$$\widehat{\theta}_{\mathbf{ab}} \pm \left[\frac{k_\alpha}{1 - k_\alpha}(\mathbf{a}'\mathbf{C}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{C}'\mathbf{a})(\mathbf{b}'\mathbf{E}\mathbf{b})\right]^{1/2}, \tag{6.57}$$

where $\widehat{\theta}_{\mathbf{ab}} = \mathbf{a}'\mathbf{\Phi}\mathbf{b}$, k_α is the $(1-\alpha)$th percentile of the largest root test statistic $T_R = e_1$ with degrees of freedom c, d, and e. In particular, for the model (6.42) with $\mathbf{C} = \mathbf{I}$ we have

$$\mathbf{a}'\mathbf{C}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{C}'\mathbf{a} = \sum a_j^2/n_j.$$

The confidence level remains valid for any number of intervals that can be deduced from $\mathbf{\Phi}$. In two important special cases, namely when $c = 1$ or

$d = 1$, κ_α can be obtained from the F distribution. This is, for instance, the case if we had planned comparing only the parameters of two particular populations only so that $c = h = 1$. Similarly, if all pairwise comparisons of only one component of the mean vectors had been planned, we can set $d = 1$. In these cases, a counterpart based on the U-statistic also leads to an F-statistic,

$$F = \frac{e_1}{h_1}\frac{1-U}{U} = \frac{e_1}{h_1}\frac{(\widehat{\theta}_{\mathbf{ab}})^2(\mathbf{aC}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{C}'\mathbf{a})^{-1}}{\mathbf{b}'\mathbf{E}\mathbf{b}} \sim F_{h_1,e_1}$$

and the $100\gamma\%$ confidence interval

$$\widehat{\theta}_{\mathbf{ab}} \pm \kappa_\gamma \left[\frac{h_1}{e_1}(\mathbf{a}'\mathbf{C}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{C}'\mathbf{a})(\mathbf{b}'\mathbf{E}\mathbf{b})\right]^{1/2}, \tag{6.58}$$

where κ_γ is the $(1-\gamma)$th quantile of the F distribution with h_1 and e_1 degrees of freedom and

$$\begin{aligned} h_1 &= h \quad \text{and} \quad e_1 = e \quad \text{if } d = 1 \\ h_1 &= d \quad \text{and} \quad e_1 = e + 1 - d \quad \text{if } h = 1. \end{aligned}$$

These intervals tend to be longer than the Bonferroni intervals unless r is large, but remains valid regardless of the number of linear combinations of interest.

Example 6.4. Comparison of anteater skulls (continued)

Consider again the data in Table 6.3. Suppose we are interested in comparing two populations. Ninety-five percent simultaneous confidence intervals for the differences in individual means of the populations could be obtained by applying (6.55) and (6.56) for the two populations are shown in the following table. The former is valid for one coefficient and for any two of the three populations and the latter applied with $r = 3$ is valid for all three coefficients for the first two populations. To ensure the confidence level for any number of intervals, we can employ (6.57). In this case, the confidence intervals of particular interest, say those for the first two populations, will be much wider than those reported below. As expected, even these less conservative intervals contain 0, thus providing no evidence to support that there is significant difference in any means.

	Scheffe Intervals		**Bonferroni Intervals**	
Mean	Lower Bound	Upper Bound	Lower Bound	Upper Bound
$\mu_{11} - \mu_{21}$	-0.04731	0.07965	-0.03286	0.06519
$\mu_{12} - \mu_{22}$	-0.05111	0.07595	-0.03665	0.06148
$\mu_{13} - \mu_{23}$	-0.07514	0.09181	-0.05613	0.07280

6.5 MANOVA WITH UNEQUAL COVARIANCES

In this section we drop the assumption of equal covariances and consider the problem of testing the equality of population means under the distributional assumption,

$$\mathbf{Y}_{ij} \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i); \quad i = 1, \ldots, I;\ j = 1, \ldots, n_i. \tag{6.59}$$

The model can also be written as

$$\mathbf{Y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\delta}_i + \boldsymbol{\epsilon}_{ij}, \quad \text{where } \boldsymbol{\epsilon}_{ij} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_i) \tag{6.60}$$

so that the null hypothesis can be expressed as

$$H_0 : \boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 \cdots = \boldsymbol{\delta}_I = \mathbf{0}. \tag{6.61}$$

This problem can be tackled by taking a modified approach that we took to solve the multivariate Behrens−Fisher problem so as to handle issues in MANOVA. There are multiple solutions to the problem, as was the case even with equal covariances, and here we present a particular class of solutions. As in the previous section, consider the transformed data $\mathbf{X}_{ij} = g(\mathbf{s})^{-1/2}\mathbf{Y}_{ij} \sim N_p(\boldsymbol{\theta}, n_i\boldsymbol{\Lambda}_i)$, where $\boldsymbol{\theta} = g(\mathbf{s})^{-1/2}\boldsymbol{\mu}$, $g(\mathbf{s})$ is any appropriate positive definite data matrix such as

$$g(\mathbf{s}) = \frac{\mathbf{s}_1}{n_1} + \frac{\mathbf{s}_2}{n_2} + \cdots + \frac{\mathbf{s}_I}{n_I}$$

or the identity matrix $\mathbf{I}_p$, and

$$\boldsymbol{\Lambda}_i = g(\mathbf{s})^{-1/2}\frac{\boldsymbol{\Sigma}_i}{n_i}g(\mathbf{s})^{-1/2}, \quad i = 1,\ 2 \ldots, I.$$

As in the case of the two sample problem, testing of the hypothesis (6.61) can be based on independent random variables

$$\overline{\mathbf{X}}_i \sim N(\boldsymbol{\theta}, \boldsymbol{\Lambda}_i), \quad \text{under } H_0 \tag{6.62}$$

and

$$\begin{aligned} \mathbf{R}_i &= (\mathbf{v}_i^{-1/2}\boldsymbol{\Lambda}_i\mathbf{v}_i^{-1/2})^{-1/2}(\mathbf{v}_i^{-1/2}\mathbf{V}_i\mathbf{v}_i^{-1/2})(\mathbf{v}_i^{-1/2}\boldsymbol{\Lambda}_i\mathbf{v}_i^{-1/2})^{-1/2} \\ &\sim W_p(n_i - 1, \mathbf{I}_p), \end{aligned} \tag{6.63}$$

where

$$\mathbf{V}_i = (n_i - 1)g(\mathbf{s})^{-1/2}\frac{\mathbf{S}_i}{n_i}g(\mathbf{s})^{-1/2} \sim W_p(n_i - 1, \boldsymbol{\Lambda}_i).$$

To establish a class of testing procedures, consider the decomposition

$$\begin{aligned} \mathbf{Z}_i &= \boldsymbol{\Lambda}_i^{-1/2}(\overline{\mathbf{X}}_i - \boldsymbol{\theta}) \\ &= \boldsymbol{\Lambda}_i^{-1/2}(\overline{\mathbf{X}}_i - \widehat{\boldsymbol{\theta}}) + \boldsymbol{\Lambda}_i^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \end{aligned} \tag{6.64}$$

where $\mathbf{Z}_i \sim N(\mathbf{0}, \mathbf{I})$ and $\widehat{\boldsymbol{\theta}}$ is the estimate (MLE when $\mathbf{\Lambda}_i$ has been specified) of the common mean vector $\boldsymbol{\theta}$, namely

$$\widehat{\boldsymbol{\theta}} = \widetilde{\overline{\mathbf{X}}} = \left(\sum_{i=1}^{I} \mathbf{\Lambda}_i^{-1} \right)^{-1} \left(\sum_{i=1}^{I} \mathbf{\Lambda}_i^{-1} \overline{\mathbf{X}}_i \right).$$

Since $\mathbf{\Lambda}_i$ are parameters that can be tackled by means of $\mathbf{R}_i$ having the observed value $\mathbf{v}_i^{1/2} \mathbf{\Lambda}_i^{-1} \mathbf{v}_i^{1/2}$, it now follows from (6.64) that

$$\begin{aligned} \mathbf{T}_i &= \mathbf{W}_i(\mathbf{\Lambda})(\overline{\mathbf{X}}_i - \widetilde{\overline{\mathbf{X}}}) \\ &= \mathbf{W_i} \mathbf{\Lambda_i^{1/2}} \mathbf{Z}_i - \mathbf{W_i} \left(\sum_{i=1}^{I} \mathbf{\Lambda}_i^{-1} \right)^{-1} \sum_{i=1}^{I} \mathbf{\Lambda}_i^{-1/2} \mathbf{Z}_i, \end{aligned} \tag{6.65}$$

$i = 1, 2, \ldots, I$ is a set of random quantities with known distributions that we can exploit to derive testing procedures, where $\mathbf{W_i} = \mathbf{W}_i(\mathbf{\Lambda})$ is a weight function of covariance matrices to be chosen appropriately to attain certain desirable properties. Two weight functions leading to solutions that reduce to familiar solutions in the univariate case and in the $I = 2$ case are

$$\mathbf{W}_i(\mathbf{\Lambda}) = \mathbf{\Lambda}_i^{-1/2} \text{ and } \mathbf{W}_i(\mathbf{\Lambda}) = \left(\sum_{i=1}^{I} \mathbf{\Lambda}_i \right) \mathbf{\Lambda}_i^{-1}.$$

Certain functions of $\mathbf{T}_i$ even have distributions free of $\mathbf{\Lambda}_i$. For example, as discussed by Gamage, Mathew, and Weerahandi (2004), if the weights are chosen as $\mathbf{W}_i(\mathbf{\Lambda}) = \mathbf{\Lambda}_i^{-1/2}$, then the random variable defined by

$$\widetilde{\mathbf{H}}_1(\mathbf{X}; \mathbf{\Lambda_1}, \ldots, \mathbf{\Lambda_I}) = \sum_{i=1}^{I} \mathbf{T}_i' \mathbf{T}_i = \sum_{i=1}^{I} (\overline{\mathbf{X}}_i - \widetilde{\overline{\mathbf{X}}})' \mathbf{\Lambda}_i^{-1} (\overline{\mathbf{X}}_i - \widetilde{\overline{\mathbf{X}}}) \tag{6.66}$$

has a chi-squared distribution with $p(I-1)$ degrees of freedom. This corresponds to the Lawley-Hotelling's test in the equal covariances case. The extreme region and the p-value based on $\widetilde{\mathbf{H}}_1$ is easily obtained as

$$\begin{aligned} p &= \Pr(\widetilde{\mathbf{H}}_1 \geq \widetilde{\mathbf{H}}_1(\mathbf{x}; \mathbf{v}_1^{1/2} \mathbf{R}_1^{-1} \mathbf{v}_1^{1/2}, \ldots, \mathbf{v}_I^{1/2} \mathbf{R}_I^{-1} \mathbf{v}_I^{1/2})) \\ &= 1 - EF_\chi(\widetilde{\mathbf{H}}_1(\mathbf{x}; \mathbf{v}_1^{1/2} \mathbf{R}_1^{-1} \mathbf{v}_1^{1/2}, \ldots, \mathbf{v}_I^{1/2} \mathbf{R}_I^{-1} \mathbf{v}_I^{1/2}), \end{aligned} \tag{6.67}$$

where F_χ is the cdf of chi-squared distribution with $p(I-1)$ degrees of freedom and the expectation is taken with respect to the Wishart random variables $\mathbf{R}_i$'s. On the other hand, the choice of weights

$$\mathbf{W}_i(\mathbf{\Lambda}) = \left(\sum_{i=1}^{I} \mathbf{\Lambda}_i \right) \mathbf{\Lambda}_i^{-1}$$

provides a solution that reduces to the test given by (6.22) in the two-sample case (see Appendix B.1). Construction of the test variable and computation of p-value with any general set of weights $\mathbf{W}_i(\mathbf{\Lambda})$ and any I is described below in a more general setting.

6.5.1 Generalized Wilks test

To establish a general procedure for computing the generalized p-value for any of the counter parts of MANOVA solutions, including the Wilks test of special interest, consider the standardized between population sum of cross products with any weight function,

$$\begin{aligned}\widetilde{\mathbf{H}}(\mathbf{X};\mathbf{\Lambda}_1,\ldots,\mathbf{\Lambda}_I) &= \sum_{i=1}^{I}\mathbf{T}_i\mathbf{T}_i' \\ &= \sum_{i=1}^{I}\mathbf{W}_i\ (\overline{\mathbf{X}}_i-\overline{\overline{\mathbf{X}}})(\overline{\mathbf{X}}_i-\overline{\overline{\mathbf{X}}})'\mathbf{W}_i \\ &= \mathbf{H}(\mathbf{Z_1}(\mathbf{\Lambda_1}),\ldots,\mathbf{Z_I}(\mathbf{\Lambda_I});\mathbf{\Lambda_1},\ldots,\mathbf{\Lambda_I}), \qquad (6.68)\end{aligned}$$

In view of the form of classical MANOVA results, we can define the potential test variable to be based on eigenvalues of the matrices

$$\mathbf{T}=\mathbf{H}(\mathbf{Z_1}(\mathbf{\Lambda_1}),\ldots,\mathbf{Z_I}(\mathbf{\Lambda_I}),\mathbf{v}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{v}_1^{1/2},\ldots,\mathbf{v}_I^{1/2}\mathbf{R}_I^{-1}\mathbf{v}_I^{1/2})$$

and

$$\mathbf{t}=\widetilde{\mathbf{H}}(\mathbf{x};\mathbf{v}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{v}_1^{1/2},\ldots,\mathbf{v}_I^{1/2}\mathbf{R}_I^{-1}\mathbf{v}_I^{1/2}),$$

where $\mathbf{x}$ is the observed value of $\mathbf{X}$. Observe that the distribution of any test variable based on these matrices is free of unknown parameters, and at the observed sample point the two matrices become equal because $(\mathbf{v}_i^{1/2}\mathbf{R}_i^{-1}\mathbf{v}_i^{1/2})_{obs}=\mathbf{\Lambda}_i$. Moreover, the eigenvalues of $\mathbf{T}$ tends to take greater values for greater deviations from the null hypothesis. Hence we could employ $(\mathbf{T},\mathbf{t})$ to define test variables. This means that any test variable based on $(\mathbf{T},\mathbf{t})$ can define extreme regions with computable probabilities and having observed sample point on its boundary. As discussed in the previous section, we can develop any counterpart of the classical testing procedures by taking the corresponding function of the eigenvalues of these matrices.

Of particular interest is the generalized Wilks test based on the p-value

$$\begin{aligned} p &= \Pr(|\mathbf{I}_p+\mathbf{T}|\geq|\mathbf{I}_p+\mathbf{t}|) \\ &= \Pr(|\mathbf{I}_p+\mathbf{H}(\mathbf{Z_1}(\mathbf{\Lambda_1}),\ldots,\mathbf{Z_I}(\mathbf{\Lambda_I}),\mathbf{v}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{v}_1^{1/2},\ldots,\mathbf{v}_I^{1/2}\mathbf{R}_I^{-1}\mathbf{v}_I^{1/2})| \\ &\geq |\mathbf{I}_p+\widetilde{\mathbf{H}}(\mathbf{x};\mathbf{v}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{v}_1^{1/2},\ldots,\mathbf{v}_I^{1/2}\mathbf{R}_I^{-1}\mathbf{v}_I^{1/2})|)\ . \qquad (6.69)\end{aligned}$$

The literature on MANOVA does not yet provide an expression for (6.69) in terms of univariate random variables. This is an area requiring further research. Nevertheless, in its current representation, this p-value or that of any other test criterion based on eigenvalues of $\mathbf{T}$ can be computed by simulating univariate standard normal random variables. This is accomplished by simulating the Wishart random matrices using independent standard normal random variates and the following result:

$$\text{If } \mathbf{Z}_j\sim N(\mathbf{0},\mathbf{I}_p),\quad j=1,\ 2,\ldots,J,\quad \text{then } \sum\mathbf{Z}_j\mathbf{Z}_j'\sim W_p(J,\mathbf{I}_p). \qquad (6.70)$$

The simulation is then carried out in the following steps:

1. Generate a large sample of random numbers from $\mathbf{Z}_i = \mathbf{\Lambda}_i^{-1/2}(\overline{\mathbf{X}}_i - \boldsymbol{\theta}) \sim N(\mathbf{0}, \mathbf{I})$ based on sets of independent standard normal random numbers.

2. Generate a large sample from each of the distributions of $\mathbf{R}_i \sim W_p(n_i - 1, \mathbf{I}_p)$ based on sets of independent standard normal random numbers.

3. For each set of simulated samples from $\mathbf{R}_i$, $i = 1,\ 2, \ldots, I$, replace $\mathbf{\Lambda}_i$ by the simulated value of $\mathbf{v}_i^{1/2}\mathbf{R}_i^{-1}\mathbf{v}_i^{1/2}$, and compute

$$\widetilde{\mathbf{H}} = \mathbf{H}_{\mathbf{s}}(\mathbf{x}; \mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_I) = \sum_{i=1}^{I} \mathbf{\Lambda}_i^{-1/2}\ (\overline{\mathbf{x}}_i - \overline{\overline{\mathbf{x}}})(\overline{\mathbf{x}}_i - \overline{\overline{\mathbf{x}}})' \mathbf{\Lambda}_i^{-1/2},$$

where $\overline{\mathbf{x}}_i = g(\mathbf{s})^{-1/2}\overline{\mathbf{y}}_i$ is computed using the actual data.

4. For each set of simulated samples from $\mathbf{R}_i, \mathbf{Z}_i$; $i = 1,\ 2, \ldots, I$, replace $\mathbf{\Lambda}_i$ by the simulated value of $\mathbf{v}_i^{1/2}\mathbf{R}_i^{-1}\mathbf{v}_i^{1/2}$, and compute

$$\mathbf{T}_i = \mathbf{Z}_i - \mathbf{\Lambda}_i^{-1/2} \left(\sum_{i=1}^{I} \mathbf{\Lambda}_i^{-1} \right)^{-1} \sum_{i=1}^{I} \mathbf{\Lambda}_i^{-1/2} \mathbf{Z}_i, i = 1,\ 2, \ldots, I$$

and then compute $\mathbf{H} = \sum_{i=1}^{I} \mathbf{T}_i \mathbf{T}_i'$,

5. Using independent sets of simulated samples generated from $\widetilde{\mathbf{H}}$ and $\mathbf{H}$, compute the fraction of pairs, $\widehat{P}$, for which $\Pr(|\mathbf{I}_p + \mathbf{H}| \geq |\mathbf{I}_p + \widetilde{\mathbf{H}}|)$.

6. Estimate the p-value by $\widehat{P}$.

Multiple comparisons can be carried out by taking an approach similar to one taken in the previous section. Simultaneous tests and confidence intervals for desired differences in mean vectors are obtained by applying the solutions to the multivariate Behrens−Fisher problem developed above with the Bonferroni adjustment.

Exercises

6.1 Consider the two multivariate populations problem with equal covariance matrices,

$$\begin{aligned} \mathbf{Y}_{1j} &\sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), \quad j = 1, 2, \cdots, m \\ \mathbf{Y}_{2j} &\sim N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}), \quad j = 1, 2, \cdots, n\ . \end{aligned} \tag{6.71}$$

Show that $\overline{\mathbf{Y}}_1$ and $\overline{\mathbf{Y}}_2$ are the MLEs of population means and that the pooled sample covariance matrix given by

$$\mathbf{V} = \frac{1}{m+n}\sum_{j=1}^{m}(\mathbf{Y}_{1j} - \overline{\mathbf{Y}}_1)(\mathbf{Y}_{1j} - \overline{\mathbf{Y}}_1)' + \sum_{j=1}^{n}(\mathbf{Y}_{2j} - \overline{\mathbf{Y}}_2)(\mathbf{Y}_{2j} - \overline{\mathbf{Y}}_2)'$$

is the MLE of the common covariance matrix. Show that all three estimates are independently distributed.

6.2 Consider model (6.38) and hypotheses of the form (6.45). Write down the form of $\mathbf{C}$ and $\mathbf{D}$ matrices if

(a) only populations 2 and 3 are to be compared with data from all populations,
(b) the hypothesis that all populations have the first coefficient in common,
(c) the hypothesis that the difference in the first coefficients of populations 2 and 3 is to be tested.

6.3 In the previous exercise, write down the particular formulae

(a) for computing p-values for each of the hypotheses,
(b) for constructing confidence regions for each quantity of interest.

6.4 Let $\mathbf{X}_{ij} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}_i)$; $i = 1, \ldots, I$; $j = 1, \cdots, n_i$ be a random sample from I multivariate populations. If covariance matrices were known, show that

$$\overline{\overline{\mathbf{X}}} = \left(\sum_{i=1}^{I}\boldsymbol{\Lambda}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\boldsymbol{\Lambda}_i^{-1}\overline{\mathbf{X}}_i\right)$$

is the MLE of the mean vector $\boldsymbol{\mu}$ and that it is unbiased.

6.5 Consider again the normal random sample in Exercise 6.4. Derive the MLE of $\boldsymbol{\Lambda}_i$ and its unbiased counterpart. Hence find the MLE $\boldsymbol{\mu}$ of when the covariances are unknown.

6.6 Consider the problem of multiple comparisons based on the data

$$\mathbf{Y}_{ij} \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i); \quad i = 1, \ldots, I;\ j = 1, \ldots, n_i.$$

Consider the problem of testing a general hypothesis involving double linear combinations of the form

$$H_0 : \mathbf{CBD} = \mathbf{0},$$

where $\mathbf{C}$ is a $c \times I$ matrix of rank $c \leq I$, and $\mathbf{D}$ is a $p \times d$ matrix of rank $d \leq p$. By taking the generalized approach or otherwise, establish procedures for testing H_0. Construct the form of confidence intervals for the parameter $\theta = \mathbf{a}'\mathbf{\Phi}\mathbf{b}$, where $\mathbf{\Phi} = \mathbf{CBD}$, $\mathbf{a}$ is a $c \times 1$ vector of constant, and $\mathbf{b}$ is a $d \times 1$ vector of constants.

6.7 Consider the data in Table 6.2.

(a) Use formulae (6.56) and (6.57) to construct 95% simultaneous intervals to be valid for all pairwise differences $\mu_{ik} - \mu_{jk}$, $k = 1, 2, 3$ of any two of the three populations.
(b) Construct 90% confidence ellipsoid for the difference in mean vectors of the first two populations.

6.8 Consider again the data in Table 6.3. Perform the following analyses without the assumption of equal covariance matrices.

(a). Compute the generalized p-value for testing equality of mean vectors of first the two populations.
(b). Construct 95% simultaneous confidence intervals for each of the differences in coefficients of the first two populations.

6.9 Consider again the data in Table 6.3. Compute the generalized p-value for testing the equality of all three mean vectors by applying (i) the generalized Lawley−Hotelling test, (ii) the generalized Wilks test.

CHAPTER 7

MIXED MODELS IN REPEATED MEASURES

7.1 INTRODUCTION

The terminology *repeated measures* or *repeated measurements*, is used to refer to data from one or more response variables, which are observed on multiple occasions.. The response variables for each experimental unit or subject are observed over time under the same or different experimental conditions. There are many real-world situations where studies are conducted for a period of time and certain measurements on the experimental subjects are taken at regular or irregular time periods. Experimental designs involving repeated measures are heavily used in clinical trials for evaluating the efficacy and long-term side effects of experimental drugs. In fact, applications of repeated measures methods can be encountered in a wide variety of fields, including biomedical research, health and life sciences, econometrics, environmetrics, industrial experiments, marketing studies, education, sociology, and psychology. The response variable that we track in such applications can range from blood pressure of a patient to the time that a worker takes to complete a certain routine job.

(Generalized Inference in Repeated Measures, Edition 2). By (Weerahandi)

Applications involving repeated measures could have one or more groups of subjects or experimental units. Most experiments in biomedical research involve a number of groups of subjects receiving different treatments and the subjects are observed periodically under the same or different conditions. The difference in experimental conditions over time might be merely due to certain uncontrolled conditions or designed experimental conditions such as the level of a factor (e.g., dose level) that is being changed from one occasion to another. From now on, for the sake of simplicity, we refer to the experimental units as subjects, regardless of the situation.

As far as the analysis of repeated measures is concerned, one special feature we need to take into account is that observations taken from a group of subjects are not independent over time as they consist of observations from the same subjects taken at different occasions. However, typically the observations from different subjects are independent of one another. Moreover, the dependence of data over time can be modeled with a few parameters. If this is not the case, we can simply analyze the data using MANOVA models with unstructured covariance matrices. Although the MANOVA approach has the advantage of milder assumptions being made, its major drawback is that the MANOVA inference procedures tend to be less efficient due to large number of unknown parameters in the unstructured covariance matrix.

In this chapter, we will study some simple models that have these features and develop procedures for making inferences concerning the underlying factors. The additional assumptions we make in the current setting will lead to structured covariance matrices with few unknown parameters and are reasonable and appropriate for applications involving repeated measurements. In this book, in addition to the variance components representing such factors as within and among subjects variations, we consider only models involving one or two factors with fixed effects. For problems and solutions involving more than two factors, the reader is referred to Vonesh and Chinchilli (1997).

There are two types of factors that we need to distinguish in analyzing repeated measurements. They are referred to as within-subject and among-subjects effects. The former includes within-subject covariates that may vary over time, time-dependent covariates, and factors that were changed during the course of the experiment, whereas the latter includes treatment groups, and among-subjects covariates such as subjects' gender and race that do not change over time.

7.2 MIXED MODELS FOR ONE GROUP

First consider the simplest possible situation in which we have just one group of subjects over time and there is only one factor, whose levels might be different over time. The levels of the factor might be the time itself or some dose levels of a treatment confounded with the time effects. Table 7.1 shows an example of a typical data set involving one group. In this data set reported by Timm (1980), each of a group of 11 subjects were given 5 probe words at different occasions and subjects mean reaction times were measured.

Table 7.1 Reaction times to probe words

Subject	P1	P2	P3	P4	P5
1	51	36	50	35	42
2	27	20	26	17	27
3	37	22	41	37	30
4	42	36	32	34	27
5	27	18	33	14	29
6	43	32	43	35	40
7	41	22	36	25	38
8	38	21	31	20	16
9	36	23	27	25	28
10	26	31	31	32	36
11	29	20	25	26	25
Means:	36.1	25.6	34.1	27.3	30.7

Consider a $T \times 1$ vector $\mathbf{Y}$ of responses or observations at T different occasions. Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iT})$ be the vector of observations taken from the ith subject. Suppose we have a sample of I subjects taken from a certain population. For the data set in Table 7.1, $T = 5$ and $I = 11$. It is not necessary that the subjects are observed at regular time intervals, but it is assumed that each subject is observed at each of the T time points. We assume that the observations of different subjects are independently distributed and that we have a complete data set from all subjects in the group. The observations at different occasions are usually dependent. Let Σ be the $T \times T$ matrix of covariances of $\mathbf{Y}$. If $T = 2$, then the problem reduces to the classical paired data problem and the significance of the difference in means of the two data columns can be tested by applying the paired t-test. If the covariance matrix Σ has no special structure, then the data can be analyzed using the multivariate methods that we studied in Chapter 6. That approach, however, will introduce too many unknown parameters into the inference problem thus making its testing procedures less efficient than they need to be. The point is that, in dealing with repeated measures, it is possible to derive the structure of the covariance matrix by making certain reasonable assumptions. Here we confine our attention to a widely used class of structured covariance matrices that can be derived by a simple mixed model.

To be specific, assume the mixed model

$$Y_{it} = \alpha_i + \beta_t + \epsilon_{it}, \tag{7.1}$$

where α_i is the random effect due to subject i, β_t, $t = 1, \ldots, T$ are fixed effects due to occasions/treatments, and ϵ_{it} are the residual terms. Assume that

$$\alpha_i \sim N(0, \sigma_\alpha^2), \qquad \epsilon_{it} \sim N(0, \sigma^2) \tag{7.2}$$

for all $t = 1, \ldots, T$; $i = 1, \ldots, I$, and that they are all independently distributed, where σ_α^2 and σ^2 are the variance components of the mixed model. Notice that this is similar to the one-way random effects model except that now we have a number of fixed effects to deal with. It is easily seen from the model (7.1) that $\text{Var}(Y_{it}) = \sigma_\alpha^2 + \sigma^2$ and that $\text{Cov}(Y_{it}, Y_{it'}) = \sigma_\alpha^2$, and hence

$$\mathbf{Y}_i \sim N_T(\boldsymbol{\beta}, \Sigma) \tag{7.3}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)$ and the covariance matrix Σ has the special structure, namely the compound symmetric structure

$$\Sigma = \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T^{'} + \sigma^2 \mathbf{I}_T, \tag{7.4}$$

where $\mathbf{1}_T$ is a $T \times 1$ vector of 1's. When the covariance matrix has such special structure, it is more convenient to derive testing procedures directly from (7.1).

7.2.1 Analysis of Variance

Define various sample means that could play a role in the analysis as

$$\begin{aligned} \overline{Y}_i &= \frac{1}{T}\sum_{t=1}^{T} Y_{it} \quad \text{for } i = 1, \ldots, I, \\ \overline{Y}_t &= \frac{1}{I}\sum_{i=1}^{I} Y_{it} \quad \text{for } t = 1, \ldots, T, \end{aligned}$$

and denote by $\overline{Y}$ the grand sample mean of all the data. An analysis of variance table providing a basis for making inferences on model (7.1) can be obtained by decomposing the total sum of squares as

$$S_T = S_A + S_B + S_E \ , \tag{7.5}$$

which is obtained by squaring and summing the identity

$$Y_{it} - \overline{Y} = (\overline{Y}_i - \overline{Y}) + (\overline{Y}_t - \overline{Y}) + (Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y}),$$

where

$$S_A = T\sum(\overline{Y}_i - \overline{Y})^2, \tag{7.6}$$

$$S_B = I\sum(\overline{Y}_t - \overline{Y})^2, \tag{7.7}$$

and

$$S_E = \sum\sum (Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y})^2. \tag{7.8}$$

Table 7.2 provides the details of the resulting ANOVA including the degrees of freedom of the above sums of squares and the expected values of the mean sums of squares, $MS = SS/DF$.

Table 7.2 One factor ANOVA

Source	DF	SS	E(MS)
Subjects	$I-1$	S_A	$T\sigma_\alpha^2 + \sigma^2$
Occasions	$T-1$	S_B	$I\sum(\beta_t - \overline{\beta})^2/(T-1) + \sigma^2$
Error	$(I-1)(T-1)$	S_E	σ^2
Total	$IT-1$	S_T	

Inferences concerning parameters of model (7.1) can now be carried out using the distributional results that the ANOVA Table 7.2 imply. First consider the problem of testing the equality of occasion means,

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_T. \tag{7.9}$$

Under H_0, we have

$$\frac{S_B}{\sigma^2} \sim \chi^2_{T-1} \tag{7.10}$$

and regardless of whether or not the null hypothesis is true, we have

$$\frac{S_E}{\sigma^2} \sim \chi^2_{(I-1)(T-1)}. \tag{7.11}$$

It is now evident that hypothesis (7.9) can be tested on the basis of the F-statistic

$$\frac{S_B/(T-1)}{S_E/(I-1)(T-1)} \sim F_{T-1,(I-1)(T-1)} \tag{7.12}$$

and the resulting p-value

$$p = 1 - H_{T-1,(I-1)(T-1)}\left(\frac{s_B/(T-1)}{s_E/(I-1)(T-1)}\right), \tag{7.13}$$

where $H_{T-1,(I-1)(T-1)}$ is the cdf of the F distribution with $T-1$ and $(I-1)(T-1)$ degrees of freedom.

Inferences about the variance components σ_α^2 and σ^2 can be performed based on (7.11) and

$$\frac{S_A}{T\sigma_\alpha^2 + \sigma^2} \sim \chi^2_{I-1}\ . \tag{7.14}$$

For example, the $100\gamma\%$ lower confidence bound for σ_α^2 is the solution σ_0^2 of the equation

$$\gamma = \int_{\frac{s_E}{s_A+s_E}}^{1} G_{T(I-1)}\left(\frac{1}{T\sigma_0^2}\left(\frac{s_A}{1-b} - \frac{s_E}{b}\right)\right) f_B(b)\, db, \qquad (7.15)$$

where $G_{T(I-1)}$ is the cdf of the chi-squared distribution with $T(I-1)$ degrees of freedom and the integration is to be performed with respect to the beta random variable, $B \sim \text{Beta}((I-1)(T-1)/2, (T-1)/2)$. The XPro software package computes the confidence limits and p-values by exact numerical integration. The integration can also be carried out using numerical integration procedures available from other statistical and mathematical software packages such as SAS and SPlus. The confidence intervals for the variance component can also be constructed by Tukey−Williams method, preferably with the Wang adjustment.

Inferences concerning the error variance, σ^2 is straightforward from (7.11). In particular,

$$\widehat{\sigma}^2 = \frac{s_E}{(I-1)(T-1)} \qquad (7.16)$$

is an unbiased estimate of σ^2.

7.2.2 Multiple comparisons

After the significance of differences between occasion means has been established, we can proceed to do multiple comparisons. Of particular interest are the pairwise comparisons. If only one such comparison had been pre-planned, the two means can be compared by a paired t-test. To see this, consider the problem of comparing occasion means β_t and $\beta_{t'}$. From (7.1) we get

$$\begin{aligned} Y_{it} &= \alpha_i + \beta_t + \epsilon_{it}, \\ Y_{it'} &= \alpha_i + \beta_{t'} + \epsilon_{it'} \end{aligned}$$

and so

$$X_i = Y_{it} - Y_{it'} = \theta + e_i \;\; i = 1, \ldots, I$$

forms a set of independent data from a normal population with mean θ and some constant variance, where $\theta = \beta_t - \beta_{t'}$ is the parameter of interest and $e_i = \epsilon_{it} - \epsilon_{it'} \sim N(0, 2\sigma^2)$ is an error term. The random variable X_i is distributed independently of $S_E/\sigma^2 \sim \chi^2_{(I-1)(T-1)}$, because

$$\sum_t X_i(Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y}) = 0.$$

Hence, inferences about θ can be based on the t-statistic

$$t = \frac{(\overline{X} - \theta)}{\widehat{\sigma}\sqrt{2/I}} \sim t_{(I-1)(T-1)}, \qquad (7.17)$$

where $\overline{X} = \overline{Y}_{.t} - \overline{Y}_{.t'} \sim N(\theta, 2\sigma^2/I)$ is the sample mean of X_i data and $S^2 = \sum(X_i - \overline{X})^2/(I-1)$ is the sample variance. For example, the $100\gamma\%$ confidence interval for $\theta = \beta_t - \beta_{t'}$ is computed using the formula

$$(\overline{y}_{it} - \overline{y}_{it'}) \pm t_{(I-1)(T-1)}(1-\alpha/2)\widehat{\sigma}\sqrt{2/I}, \tag{7.18}$$

where $t_{(I-1)(T-1)}(k)$ is the kth quantile of the Student's t distribution with $(I-1)(T-1)$ degrees of freedom and $\alpha = 1-\gamma$. If there were r prespecified pairwise comparisons of interest, the simplest way to obtain simultaneous confidence intervals is to apply the Bonferroni method. Then, for the values of t and t' of interest we apply the formula

$$(\overline{y}_{it} - \overline{y}_{it'}) \pm t_{(I-1)(T-1)}(1-\alpha/2r)\widehat{\sigma}\sqrt{2/I}. \tag{7.19}$$

As in other problems of multiple comparisons, the conservative intervals obtained in this manner tend to be shorter than those given by more complicated methods such as the Scheffe method.

Example 7.1. Comparison of mean reaction times to probe words

Consider the data in Table 7.1. The table also provides the sample mean reaction times, most of which seem significantly different. To establish the statistical significance and to detect the ones that are different, let us carry out the above analysis. The sums of squares and the F-values are summarized in the following ANOVA table.

Source	DF	SS	F-value
Subjects	10	1991	8.489
Occasions	4	868	9.248
Error	40	938	
Total	54	3796	

The p-value for testing the hypothesis of equal mean reaction times is $p = 1 - H_{4,40}(9.248) = 0.0$ suggesting that the rejection of the null hypothesis. Therefore, we can proceed to make pairwise comparisons. To apply formula (7.19) to construct simultaneous 95% confidence for all possible pairs, we set $r = 10$. The estimated error standard deviation is $\widehat{\sigma}_e = \sqrt{938/40} = 4.843$ and so, the half width of each confidence interval is computed as

$$\begin{aligned} H_w &= t_{40}(.9975)4.843\sqrt{2/11} \\ &= 2.971 * 2.065 = 6.14. \end{aligned}$$

It is now evident that the differences of mean reaction times to probe words (1,2), (1,4), (2,3), and (3,4) are significant at the 0.05 level. The 95% equal-tail confidence interval for the among-subject variance, σ_α^2, computed using formula (7.15) is [14.35, 117.7], with its unbiased estimate at 35.7.

7.3 ANALYSIS OF DATA FROM TWO FACTORS

Now consider the case of a number of groups, which we refer to as treatment groups for simplicity of terminology. The treatments represent the levels of one factor and, as in the previous section, occasions represent another. In addition to these factors, among-subject variation measured by a variance component is also important. The design and the mixed model we consider in this section is the most widely used extension of the one treated in the previous section. In this extension, we simply have a number of groups of subjects, each of which are to be modeled as before. There is only one measure or response variable of importance, whose measurements are taken from each subject at each of the occasions. Now we have the additional problem of comparing the treatment groups and addressing the possibility that treatment groups might have an interaction with occasions or the second factor.

Suppose there are G groups of subjects on which we have repeated measures on the response variable. Suppose there are n_g subjects in group g and let $\sum n_g = N$ be the total number of subjects used in the experiment. Each subject is observed at T equally or unequally spaced time points. Let $Y_{i(g)t}$ denote the observation taken at the tth time point from the ith subject in group g. Table 7.4 below shows a typical data set in which $G = 3$ and $n_g = 7$ for all groups. This type of data can be analyzed by MANOVA methods in Chapter 6 if we do not make any assumptions about the structure of the covariance matrix. That approach, however, does not take advantage of the special covariance structures that occur due to the repeated measurements taken from subjects and thus leaves too many unknown parameters. In the other extreme, if data taken over time are nearly independent, say after controlling the within-subject variation with some historical data from each subject, one can simply use conventional ANOVA and regression methods to make all kinds of inferences.

In many applications of repeated measures, however, neither of the above two extreme solutions might be satisfactory. In such situations, by modeling the repeated measurements appropriately we can derive models with few nuisance parameters. Here we discuss a particular mixed model that is widely used and studied in the literature.

Extending model 7.1 to the case of G groups, we formulate the mixed model

$$Y_{i(g)t} = \theta_g + \beta_t + \gamma_{gt} + \alpha_{i(g)} + \epsilon_{i(g)t}, \tag{7.20}$$

for $t = 1, \ldots, T$; $i(g) = 1, \ldots, n_g$; $g = 1, \ldots, G$, where $\alpha_{i(g)}$ is the random effect due to among-subject variation, θ_g, $g = 1, \ldots, G$ are the treatment (or factor 1) effects β_t, $t = 1, \ldots, T$ are effects due to occasions (or factor 2), γ_{gt} are their interactions, and ϵ_{it} are the residual terms. Extending the usual assumption about variance components and assuming equal error variances, we have

$$\alpha_{i(g)} \sim N(0, \sigma_\alpha^2),\ \epsilon_{i(g)t} \sim N(0, \sigma^2); \tag{7.21}$$

for all $t = 1, \ldots, T$; $i(g) = 1, \ldots, n_g; g = 1, \ldots, G$.

In designed experiments involving comparison of treatment groups based on repeated measures, among-subjects variation entering into the model can be substantially minimized, if some observations are taken over time from each subject before the treatments are administered. Let $X_{i(g)t}$ be the original data from such an experiment. Suppose from each subject we have T' observations taken at time points $t' = 1, \ldots, T'$ before the treatments are given and we have T observations taken at time points $t = 1, \ldots, T$ after the treatments are given. Then, before we use model (7.20) we could transform the data by an appropriate transformation such as

$$Y_{i(g)t} = X_{i(g)t} - \overline{X}_{i(g)}, \tag{7.22}$$

where

$$\overline{X}_{i(g)} = \frac{\sum_{t'=1}^{T'} X_{i(g)t'}}{T'}.$$

To illustrate the importance of appropriate transformation of data before we use model (7.20), suppose a market analyst has data from a price trial on a consumer durable. Suppose the analyst has longitudinal data from a sample of stores selling the product for a number of time periods before the price trial and after the price trial. The sample consists of some *control stores* where the price is not increased as well as from the *test stores* where the price was increased during the price trial. In this type of applications, there could be a substantial variation in among-store sales depending on the size and type of stores. Nevertheless, the problem can be easily tackled by transformation of the data before we apply model (7.20). In this application, however, the appropriate transformation might be $Y_{i(g)t} = \log(X_{i(g)t}) - \overline{lX}$ rather than (7.22), because the *price elasticity of demand* or the percent decrease in sales is the quantity that tends to be constant among stores of different size, where $\overline{lX}$ is the sample mean of $\log(X_{i(g)t})$ data before the price increase. Moreover, in some applications involving trends in sales, one may want to transform the data as $Y_{i(g)t} = \log(X_{i(g)t}) - c\,\overline{lX}$, where c is a constant that might have been estimated by regression methods. We will address this issue further later in this chapter. Also, for further details of this type of problems and related issues, the reader is referred to Koschat and Weerahandi (2003).

In the following treatment we assume that, when historical data are available, the original data have already been appropriately transformed and that model (7.20) is appropriate for the transformed data. All time-dependent parameters can be represented in a single equation by rewriting model (7.20) in terms of the vector of all observations from subject $i(g)$,

$$\mathbf{Y}_{i(g)} = \left(Y_{i(g)1}, Y_{i(g)2}, .., Y_{i(g)T}\right)'.$$

Then, we can define our model as

$$\mathbf{Y}_{i(g)} = \mu \mathbf{1}_T + \delta_g \mathbf{1}_T + \boldsymbol{\beta} + \boldsymbol{\gamma}_g + \boldsymbol{\varepsilon}_{i(g)} \ , \tag{7.23}$$

where $\theta_g = \mu + \delta_g$, μ is the grand mean so that $\sum \delta_g = 0$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_g$ are $T \times 1$ vectors formed by β_t's and γ_{gt}'s, and

$$\boldsymbol{\varepsilon}_{i(g)} \sim N(\mathbf{0},\Sigma), \quad \text{with } \Sigma = \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T^{'} + \sigma^2 \mathbf{I}_T.$$

In some treatments of the problem found in the literature, another variance component is introduced for the interaction between the subjects and the occasions. A major drawback with such models is that they leave one variance component inestimable, thus indicating over parameterization of the problem. However there are other covariance structures under which the parameters are identifiable. For a discussion on such covariance structures and for approximate results the reader is referred to Vonesh and Chinchilli (1997).

First consider the problem of estimating unknown parameters of the model. This will also prove to be important in developing the ANOVA based on an orthogonal decomposition of the total sums of squares. Point estimation of location parameters requires some constraints to make them estimable. We impose the following natural constraints, which are important in developing tractable distributions as well:

$$\sum_{g=1}^{G} (n_g \boldsymbol{\delta}_g) = 0, \quad \mathbf{1}_T^{'} \boldsymbol{\beta} = 0, \quad \mathbf{1}_T^{'} \boldsymbol{\gamma}_g = 0 \ \forall g, \tag{7.24}$$

and

$$\sum_{g=1}^{G} \left(n_g \boldsymbol{\gamma}_g \right) = \mathbf{0}.$$

Let

$$\overline{Y}_{gt} = \frac{1}{n_g} \sum_{i(g)}^{n_g} Y_{i(g)t},$$

$g = 1, \ldots, G$, $t = 1, \ldots, T$ be the sample mean of n_g observations from the subjects in group g and let $\overline{Y}$ be the grand mean of all the data. Also define the $T \times 1$ vector

$$\overline{\mathbf{Y}}_g = \begin{pmatrix} \overline{Y}_{g1} \\ \overline{Y}_{g2} \\ \vdots \\ \overline{Y}_{gT} \end{pmatrix},$$

and define various sample means, namely the group means, occasion means, and subject means, as

$$\overline{Y}_g = \frac{1}{n_g T} \sum_{i(g)}^{n_g} \sum_{t=1}^{T} Y_{i(g)t} = \frac{\mathbf{1}_T^{'} \overline{\mathbf{Y}}_g}{T} \ ,$$

$$\overline{Y}_t = \frac{1}{N}\sum_{g=1}^{G}\sum_{i(g)}^{n_g} Y_{i(g)t},$$
$$\overline{Y}_{i(g)} = \frac{1}{T}\sum_{t=1}^{T} Y_{i(g)t} \ ,$$

where $N = \sum_{g=1}^{G} n_g$. The maximum likelihood estimates of the parameters under the constraints (7.24) are

$$\widehat{\mu} = \overline{Y}, \tag{7.25}$$
$$\widehat{\delta}_g = \overline{Y}_g - \overline{Y}, \tag{7.26}$$
$$\hat{\boldsymbol{\beta}} = \sum_{g=1}^{G} \left(n_g \left(\overline{\mathbf{Y}}_g - \overline{Y}_g \mathbf{1}\right)\right)/N, \tag{7.27}$$

and

$$\widehat{\boldsymbol{\gamma}}_g = \overline{\mathbf{Y}}_g - \overline{Y}_g \mathbf{1} - \hat{\boldsymbol{\beta}}, \tag{7.28}$$

The point estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_g$ also take the familiar and intuitive forms when their components are written as

$$\beta_t = \overline{Y}_t - \overline{Y}, \tag{7.29}$$
$$\gamma_{gt} = \overline{Y}_{gt} - \overline{Y}_g - \overline{Y}_t + \overline{Y} \tag{7.30}$$

7.4 ANOVA UNDER EQUAL ERROR VARIANCES

In this section we develop the classical ANOVA under the common assumption of equal error variances, an assumption made in almost all treatments of the problem found in the literature. In the next chapter we shall relax this assumption, which is usually made for mathematical tractability and simplicity.

In view of the form of the point estimates of parameters in the equal variances case given above, consider the decomposition of the sums of deviations of sample means,

$$\overline{Y}_{i(g)} - \overline{Y} = (\overline{Y}_g - \overline{Y}) + (\overline{Y}_{i(g)} - \overline{Y}_g) \tag{7.31}$$

and

$$\begin{aligned}(Y_{i(g)t} - \overline{Y}) = {} & (\overline{Y}_{i(g)} - \overline{Y}) + (\overline{Y}_t - \overline{Y}) + \\ & (\overline{Y}_{gt} - \overline{Y}_t - \overline{Y}_g + \overline{Y}) + \\ & (Y_{i(g)t} - \overline{Y}_{gt} - \overline{Y}_{i(g)} + \overline{Y}_g).\end{aligned} \tag{7.32}$$

The sum of cross products of any two terms on the right-hand side of (7.31) and (7.32) is zero, implying that the corresponding vectors are orthogonal. For example,

$$\begin{aligned} & \sum_{t=1}^{T}\sum_{g=1}^{G}\sum_{i(g)}^{n_g}(\overline{Y}_t - \overline{Y})(Y_{i(g)t} - \overline{Y}_{gt} - \overline{Y}_{i(g)} + \overline{Y}_g) \\ = & \sum_{t=1}^{T}\sum_{g=1}^{G}(\overline{Y}_t - \overline{Y})\sum_{i(g)}^{n_g}(Y_{i(g)t} - \overline{Y}_{gt} - \overline{Y}_{i(g)} + \overline{Y}_g) \\ = & \sum_{t=1}^{T}\sum_{g=1}^{G}n_g(\overline{Y}_t - \overline{Y})(\overline{Y}_{gt} - \overline{Y}_{gt} - \overline{Y}_g + \overline{Y}_g) = 0. \end{aligned}$$

Hence, the summation of squared terms in (7.31) and (7.32) yield the orthogonal decomposition of sums of squares,

$$S_t = S_g + S_{wg} + S_o + S_{og} + S_e, \tag{7.33}$$

where

$$S_o = N\sum_{t=1}^{T}(\overline{Y}_t - \overline{Y})^2, \tag{7.34}$$

$$S_g = T\sum_{g=1}^{G}n_g(\overline{Y}_g - \overline{Y})^2, \tag{7.35}$$

$$S_{og} = \sum_{t=1}^{T}\sum_{g=1}^{G}n_g(\overline{Y}_{gt} - \overline{Y}_t - \overline{Y}_g + \overline{Y})^2, \tag{7.36}$$

$$S_{wg} = T\sum_{g=1}^{G}\sum_{i\in g}(\overline{Y}_{i(g)} - \overline{Y}_g)^2, \tag{7.37}$$

$$S_e = \sum_{t=1}^{T}\sum_{g=1}^{G}\sum_{i(g)}^{n_g}(Y_{i(g)t} - \overline{Y}_{gt} - \overline{Y}_{i(g)} + \overline{Y}_g)^2, \tag{7.38}$$

and

$$S_t = \sum_{t=1}^{T}\sum_{g=1}^{G}\sum_{i(g)}^{n_g}(Y_{i(g)t} - \overline{Y})^2. \tag{7.39}$$

Due to the orthogonality of the vectors on which they are based, the sums of squares on the right hand side of (7.33) are independently distributed. The distribution of each sum of squares can be easily derived by averaging (7.20) appropriately and using (7.21). For example, it follows from

$$\begin{aligned} \overline{Y}_{i(g)} &= \mu + \delta_g + \alpha_{i(g)} + \overline{\epsilon}_{i(g)} \\ &\sim N(\mu + \delta_g, \sigma^2_\alpha + \sigma^2/T) \quad \forall \;\; i = 1, \dots, \; n_g \end{aligned} \tag{7.40}$$

that

$$W_g = \sum_{i \in g} \frac{(\overline{Y}_{i(g)} - \overline{Y}_g)^2}{\sigma^2_\alpha + \sigma^2/T} \sim \chi^2_{n_g - 1} \tag{7.41}$$

and that the random variables $W_1, \dots, W_{n_g}$ are independently distributed. Hence, we get

$$\frac{S_{wg}}{T\sigma^2_\alpha + \sigma^2} \sim \chi^2_{N-G} \; . \tag{7.42}$$

By expressing each term in the decomposition (7.33), in terms of its parameters and random effects, it is shown similarly that the distribution of each sum of squares term appearing in (7.33) divided by the expected mean sum of squares has a chi-squared distribution. Their degrees of freedom and the expected value of the mean sum of squares are summarized in the ANOVA Table 7.3. As we will show below, various tests on main effects and interactions can be performed based on quantities from the ANOVA table.

Table 7.3 Two-factor repeated measures ANOVA: Expected values

Source	DF	SS	E(MS)
Groups (factor 1)	$G-1$	S_g	$T\sum n_g {\delta_g}^2/(G-1) + \sigma^2_w$
Within group (subjects)	$N-G$	S_{wg}	$\sigma^2_w = T{\sigma^2}_\alpha + \sigma^2$
Occasions (factor 2)	$T-1$	S_o	$N\sum {\beta_t}^2/(T-1) + \sigma^2$
Groups × Occasions	$(T-1)(G-1)$	S_{og}	$\frac{\sum\sum n_g \gamma^2_{gt}}{(T-1)(G-1)} + \sigma^2$
Error	$(N-G)(T-1)$	S_e	σ^2
Total	$NT-1$	S_t	

7.4.1 Testing the fixed effects

Comparison of fixed effects as well inferences concerning the variance components can be based on Table 7.3. First consider the problem of comparing the main effects due to the treatments. The hypothesis that all group means are equal, namely that $\theta_1 = \theta_2 = \cdots = \theta_T$, is equivalent to the hypothesis

$$H_{01} : \delta_1 = \delta_2 = \cdots = \delta_T = 0,$$

and so it can be tested based on the F-statistic

$$F_1 = \frac{\text{MSG}}{\text{MSWG}} = \frac{S_g/(G-1)}{S_{wg}/(N-G)} \sim F_{G-1,N-G}. \tag{7.43}$$

The hypothesis is rejected for large values of the F-statistic. The p-value of the test is

$$p = 1 - H_{G-1,N-G}\left(\frac{s_g/(G-1)}{s_{wg}/(N-G)}\right), \tag{7.44}$$

where $H_{G-1,N-G}$ is the cdf of the F distribution with $G-1$ and $N-G$ degrees of freedom. Similarly, tests of the hypothesis that the occasion means are equal, i.e.,

$$H_{02} : \beta_1 = \beta_2 = \cdots = \beta_T = 0$$

can be tested on the basis of the F-statistic

$$F_2 = \frac{\text{MSO}}{\text{MSE}} = \frac{S_o/(T-1)}{S_e/(N-G)(T-1)} \sim F_{T-1,(N-G)(T-1)}. \tag{7.45}$$

Finally, consider the equality of interaction terms

$$H_{02} : \gamma_{11} = \gamma_{12} = \cdots = \gamma_{GT} = 0,$$

which means that treatment group mean profiles are parallel. Obviously, we can test H_{02} using the F-statistic

$$F_3 = \frac{\text{MSOG}}{\text{MSE}} = \frac{S_{og}/(T-1)(G-1)}{S_e/(N-G)(T-1)} \sim F_{(T-1)(G-1),(N-G)(T-1)}. \tag{7.46}$$

These results can be summarized in an ANOVA table, which is sometimes known as Repeated Measures ANOVA and is abbreviated as RM ANOVA . Notation used in the column heads are the same as that of conventional ANOVA. A number of software packages such as SAS, SPSS, SPlus, and XPro have procedures for performing various ANOVA methods with repeated measures including the RM ANOVA presented in this section. With the SAS package PROC GLM needs to be applied with the REPEATED statement. XPro has menu-driven tools for alternative models requiring no programming.

Two-factor RM ANOVA: F-values

Source	DF	SS	MS	F-Value
Groups	$G-1$	S_g	$MS_g = \frac{S_g}{G-1}$	$\frac{MS_g}{MS_{wg}}$
Within group	$N-G$	S_{wg}	$MS_{wg} = \frac{S_{wg}}{N-G}$	
Occasions	$T-1$	S_o	$MS_o = \frac{S_o}{T-1}$	$\frac{MS_o}{MS_e}$
Groups × Occasions	$(G-1)(T-1)$	S_{og}	$MS_{og} = \frac{S_{og}}{(G-1)(T-1)}$	$\frac{MS_{og}}{MS_e}$
Error	$(N-G)(T-1)$	S_e	$MS_e = \frac{S_e}{(N-G)(T-1)}$	
Total	$NT-1$	S_t		

7.4.2 Testing the variance components

Inferences about the variance components σ_α^2 and σ^2 are performed based on the distributional results

$$\frac{S_{wg}}{T\sigma_\alpha^2 + \sigma^2} \sim \chi^2_{N-G}, \tag{7.47}$$

and

$$V = \frac{S_e}{\sigma^2} \sim \chi^2_{(N-G)(T-1)}. \tag{7.48}$$

Inference on the error variance is straightforward from (7.48). In particular, the unbiased estimate of σ^2 is

$$\widehat{\sigma}^2 = \frac{s_e}{(N-G)(T-1)}.$$

Similarly, the unbiased estimate of the among-subject variance based on the above results is

$$\widehat{\sigma}_\alpha^2 = \frac{s_{wg}/(N-G) - s_e/(N-G)(T-1)}{T}.$$

As we discussed in Chapter 3, this type of estimate can frequently become negative, a well known undesirable property of the unbiased estimates of variance components.

Other inferences on the variance component follow from the results in Chapter 3. In the current application, the generalized p-value for testing the null hypothesis

$$H_0 : \sigma_\alpha^2 \leq \sigma_0^2$$

is obtained as

$$p = 1 - EG_{N-G}\left(\frac{s_{wg}}{T\sigma_0^2 + s_e/V}\right), \tag{7.49}$$

where G_{N-G} is the cdf of the chi-squared distribution with $N-G$ degrees of freedom and the and expectation is taken with respect to the random variable V defined by (7.48). Moreover, the generalized $100\gamma\%$ lower confidence bound for σ_α^2 is the solution σ_0^2 of the equation,

$$\gamma = \int_{\frac{s_e}{s_{wg}+s_e}}^{1} G_{T(N-G)}\left(\frac{1}{T\sigma_0^2}\left(\frac{s_{wg}}{1-b} - \frac{s_e}{b}\right)\right) f_B(b)\, db, \tag{7.50}$$

where $G_{T(N-G)}$ is the cdf of the chi-squared distribution with $T(N-G)$ degrees of freedom and the integration is to be performed with respect to the beta random variable $B \sim \text{Beta}((N-G)(T-1)/2, (N-G)/2)$. In this situation also the confidence intervals for the variance component can be constructed by the Tukey$-$Williams method, preferably with the Wang adjustment. The XPro software package computes both sets of confidence intervals and p-values by exact numerical integration. It also provides p-values for testing each of the three hypotheses on fixed effects of the model. Inferences about the error variance, σ_e^2 is straightforward from (7.48).

7.4.3 Multiple comparisons

Unlike in regular ANOVA, in repeated measures ANOVA, the problem of multiple comparisons is not an easy task. In fact, the arguments in specialized approaches such as the Tukey method and the Scheffe method do not go through in the current problem, except in some special cases. Perhaps the best and the easiest way to perform multiple comparisons in repeated measures is using pairwise comparisons with the Bonferroni size adjustment. To compare two treatment groups, first deduce the distribution of the estimate of group means as

$$\begin{aligned} \widehat{\theta}_g &= \overline{Y}_g = \theta_g + \overline{\alpha}_g + \overline{\epsilon}_g \\ &\sim N\left(\theta_g, \frac{1}{Tn_g}(T\sigma_\alpha^2 + \sigma^2)\right). \end{aligned} \tag{7.51}$$

Let $g1$ and $g2$ be the two groups of interest. Then we get

$$\overline{Y}_{g1} - \overline{Y}_{g2} \sim N\left(\theta_{g1} - \theta_{g2}, \frac{1}{T}(T\sigma_\alpha^2 + \sigma^2)(\frac{1}{n_{g1}} + \frac{1}{n_{g2}})\right),$$

and then using (7.42) we get

$$\frac{(\overline{Y}_{g1} - \overline{Y}_{g2}) - (\theta_{g1} - \theta_{g2})}{\sqrt{\frac{1}{T}(\frac{1}{n_{g1}} + \frac{1}{n_{g2}})S_{wg}/(N-G)}} \sim t_{N-G}. \tag{7.52}$$

The two groups $g1$ and $g2$ can now be compared based on (7.52). In particular the $100\gamma\%$ confidence interval for $\theta = \theta_{g1} - \theta_{g2}$ is computed using the formula

$$(\overline{y}_{g1} - \overline{y}_{g2}) \pm t_{(N-G)}(1 - \frac{\alpha}{2})\sqrt{\frac{1}{T}(\frac{1}{n_{g1}} + \frac{1}{n_{g2}})\left(\frac{s_{wg}}{N-G}\right)}, \tag{7.53}$$

where $t_{N-G}(k)$ is the kth quantile of the Student's t distribution with $N-G$ degrees of freedom and $\alpha = 1 - \gamma$.

If there were r prespecified pairwise comparisons of interest, simultaneous confidence intervals can be obtained by applying (7.53) with the Bonferroni adjustment. For example, in constructing simultaneous generalized confidence intervals for r pairs of differences in group means, we apply (7.53) with α/r in place of $\alpha = 1-\gamma$. Therefore, the $100\gamma\%$ simultaneous interval for a particular pair, say $\theta = \theta_{g_1} - \theta_{g_2}$, is computed as

$$(\overline{y}_{g1} - \overline{y}_{g2}) \pm t_{(N-G)}(1 - \frac{\alpha}{2r})\sqrt{\frac{1}{T}(\frac{1}{n_{g1}} + \frac{1}{n_{g2}})\left(\frac{s_{wg}}{N-G}\right)},$$

Example 7.2. Comparison of diet supplements

Consider the data set shown in Table 7.4, an example on repeated measures reported by Crowder and Hand (1990). In this example, the response variable was the effect (measured in terms of weights) of a vitamin E diet supplement on growth of guinea pigs. The body weights of the animals were recorded at the end of the weeks 1, 3, 4, 5, 6, and 7. The guinea pigs were given a growth-inhibiting substance during week 1, and then the vitamin E therapy was started at the beginning of week 5. Three groups of guinea pigs, receiving zero, low, and high doses of vitamin E, were administered during the 6-weeks period.

Table 7.4 Effect of diet supplement on growth rates

	Week:	1	3	4		5	6	7	
	Time:	t'_1	t'_2	t'_3		t_1	t_2	t_3	
Group	**Subject**	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\overline{X}_i(g)$	$\mathbf{X}_4$	$\mathbf{X}_5$	$\mathbf{X}_6$	$\overline{X}_{i(g)}$
1	1	455	460	510	475.0	504	436	466	468.7
1	2	467	565	610	547.3	596	542	587	575.0
1	3	445	530	580	518.3	597	582	619	599.3
1	4	485	542	594	540.3	583	611	612	602.0
1	5	480	500	550	510.0	528	562	576	555.3
2	6	514	560	565	546.3	524	552	597	557.7
2	7	440	480	536	485.3	484	567	569	540.0
2	8	495	570	569	544.7	585	576	677	612.7
2	9	520	590	610	573.3	637	671	702	670.0
2	10	503	555	591	549.7	605	649	675	643.0
3	11	496	560	622	559.3	622	632	670	641.3
3	12	498	540	589	542.3	557	568	609	578.0
3	13	478	510	568	518.7	555	576	605	578.7
3	14	545	565	580	563.3	601	633	649	627.7
3	15	472	498	540	503.3	524	532	583	546.3

The subject means, $\overline{X}_{i(g)}$, before and after the start of vitamin E therapy are also shown in Table 7.4. Figure 7.1 illustrates the profiles of individual guinea pigs belonging to each of the three groups. The figure does not suggest that the treatment groups are significantly different. It also indicates that most of the guinea pigs have been growing over time and hence perhaps it is more appropriate to analyze the data by growth curves methods that we will study later in this book. In any case the treatment groups can be analyzed by RM ANOVA without assuming any parametric model for the growth curves.

One may wish to apply model (7.20) to raw data from all six weeks. However, if comparison of the three doses of vitamin E is the task of primary importance, then the raw data should be transformed using formula (7.22).

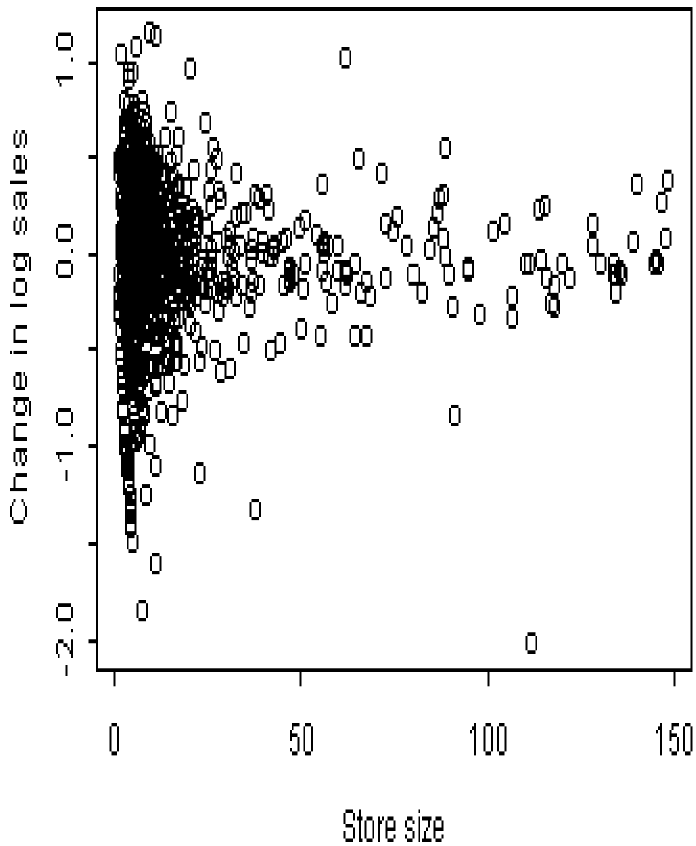

Figure 7.1 Subject profile plots by group

Then, we could base our analysis on transformed data given in the table below.

Diet supplement on growth rates: Transformed data

	Time:	t_1	t_2	t_3
Group	**Subject\Data:**	$\mathbf{Y}_4$	$\mathbf{Y}_5$	$\mathbf{Y}_6$
1	1	29.0	-39.0	-9.0
1	2	48.7	-5.3	39.7
1	3	78.7	63.7	100.7
1	4	42.7	70.7	71.7
1	5	18.0	52.0	66.0
2	6	-22.3	5.7	50.7
2	7	-1.3	81.7	83.7
2	8	40.3	31.3	132.3
2	9	63.7	97.7	128.7
2	10	55.3	99.3	125.3
3	11	62.7	72.7	110.7
3	12	14.7	25.7	66.7
3	13	36.3	57.3	86.3
3	14	37.7	69.7	85.7
3	15	20.7	28.7	79.7

Various sums of squares needed in constructing the ANOVA for testing fixed effects can be computed by applying equations (7.34)−(7.39). They computations can be conveniently carried out using software packages such as SAS, SPlus, SPSS, and XPro; XPro provides exact inference on variance components as well. ANOVA tables obtained by applying transformed data as well as the raw data are shown below.

ANOVA with raw data

Source	DF	SS	F-value	p-value
Treatments	2	18,548	1.05	0.38
Within group	12	105,434		
Weeks	5	142,555	52.55	0.00
Treatments × Weeks	10	9,763	1.80	0.08
Error	60	32,553		
Total	89	308,852		

ANOVA with transformed data

Source	DF	SS	F-value	p-value
Treatments	2	4,075	0.77	.49
Within group	12	31,797		
Weeks	2	17,192	23.83	0.00
Treatments × Weeks	4	6,175	4.28	0.01
Error	24	8,657		
Total	44	67,895		

Clearly, regardless of whether we transform the data or not, the data do not provide sufficient evidence to suspect that there is any difference between the treatment groups. The week effects are of course highly significant, a finding of less importance, because the guinea pigs were growing during the experiment. Raw data provides some week evidence of interaction between the effects of the treatments and weeks. However, with transformed data we can conclude that there is significant interaction between the treatments and the weeks, suggesting further investigation on the effects of the treatments. If we had sufficient data to detect the statistical significance of the difference in treatments, then the effects of the treatments would not have been the same over the weeks. The unbiased estimate of the error variance computed using the raw data is $\widehat{\sigma}^2 = 32553/60 = 542.6$ and the 95% confidence interval computed using the F-statistic is [390.8, 804.1]. The unbiased estimate of the among-subject variance is $\widehat{\sigma}^2_\alpha = (105434/12 - 32553/60)/5 = 774$. Its confidence intervals can be computed by applying formula (7.50). In particular, the 95% equal-tail interval for σ^2_α is [658.5 3896.9].

7.5 OTHER TWO-FACTOR MODELS

In the layout used in model (7.20), subjects were nested under the levels of one factor, say treatment groups. Measurements from the same subject were not taken under the levels of both factors. Now consider the two-factor model when the subjects are not nested under the levels of one factor but rather, observations from each subject is taken at levels of both factors. Let A and B denote the two factors and let Y_{iab} denote the measurement taken from ith subject when the level of factor A takes on the value a and that of B takes on the value b. There are two widely used designs, which are appropriate and practical depending on the experiment and the way data become available.

7.5.1 Cross classified design

For the sake of simplicity of notation, let A and B also denote the number of levels of factors A and B, respectively. Suppose it is possible to obtain data from subjects for each pair of factor level (A_a, B_b); $a = 1, \ldots, A$ and

$b = 1, \ldots, B$. When one of the factors, say factor B, is time/occasion itself, as is typically the case, this is practical if, at each occasion, data from the subjects can be obtained for each of the A levels of factor A. For example, consider a group of students undergoing an experiment, say to study the effects of tutoring before, during, and after the tutoring period. In this application indeed one can obtain test scores of students from a number of subjects for the same marking period. This is also the case if a number of related measures are taken at each occasion from a number of patients. When we have a complete set of data according to a cross-classified design, the data can be set out as in Table 7.5.

Table 7.5 Layout for the cross classified design

Subject		B_1	B_2	$\cdots$	B_B
1	A_1	Y_{111}	Y_{112}	$\cdots$	Y_{11B}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	A_1	Y_{I11}	Y_{I12}	$\cdots$	Y_{I1B}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	A_a	Y_{1a1}	Y_{1a2}	$\cdots$	Y_{1aB}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	A_a	Y_{Ia1}	Y_{Ia2}	$\cdots$	Y_{IaB}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	A_A	Y_{1A1}	Y_{1A2}	$\cdots$	Y_{1AB}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	A_A	Y_{IA1}	Y_{IA2}	$\cdots$	Y_{IAB}

Noting that this is a balanced three-way factorial design of the factors A, B, and subjects, with no replications, and distinguishing the fixed effects and random effect terms, we can formulate the mixed model

$$\begin{aligned} Y_{iab} &= \mu + \alpha_a + \beta_b + \gamma_{ab} + \iota_i + \varepsilon_{ia} + \epsilon_{ib} + e_{iab}, \\ a &= 1, \ldots, A; b = 1, \ldots, B;\ i = 1, \ldots,\ I, \end{aligned} \tag{7.54}$$

where ι_i, ε_{ia}, ϵ_{ib} are random effects due to subjects, and their interactions between the two factors, α_a, β_b, and γ_{ab}, are fixed effects of factors A, B, and their interaction, and e_{iab} are the residual terms. In order to make the parameters identifiable, we assume that, except for the individual random effect ι_i, the error term e_{iab} (both of which are drawn from a population) and the grand mean μ, all other fixed effects and interaction random effects are normalized to sum to 0 when summed over a or b. We also make the usual assumptions

$$\iota_i \sim N(0, \sigma_\iota^2), \quad \varepsilon_{ia} \sim N(0, \sigma_{\iota a}^2), \quad \epsilon_{ib} \sim N(0, \sigma_{\iota b}^2), \tag{7.55}$$

and

$$e_{iab} \sim N(0, \sigma^2) \tag{7.56}$$

and that they are independently distributed. In this case the fixed effects are easily estimated as

$$\begin{aligned} \widehat{\mu} &= \overline{Y},\ \widehat{\alpha}_a = \overline{Y}_a - \overline{Y},\ \widehat{\beta}_b = \overline{Y}_b - \overline{Y} & (7.57) \\ \gamma_{ab} &= \overline{Y}_{ab} - \overline{Y}_a - \overline{Y}_b + \overline{Y} & (7.58) \end{aligned}$$

and the estimates of the random effects would follow from the ANOVA table, which is derived from the orthogonal decomposition

$$\begin{aligned} (Y_{iab} - \overline{Y}) &= (\overline{Y}_a - \overline{Y}) + (\overline{Y}_b - \overline{Y}) + (\overline{Y}_i - \overline{Y}) + (\overline{Y}_{ab} - \overline{Y}_a - \overline{Y}_b + \overline{Y}) \\ &\quad + (\overline{Y}_{ia} - \overline{Y}_i - \overline{Y}_a + \overline{Y}) + (\overline{Y}_{ib} - \overline{Y}_i - \overline{Y}_b + \overline{Y}) \\ &\quad + (Y_{iab} - \overline{Y}_{ab} - \overline{Y}_{ia} - \overline{Y}_{ib} + \overline{Y}_i + \overline{Y}_a + \overline{Y}_b - \overline{Y}), \qquad (7.59) \end{aligned}$$

where

$$\begin{aligned} \overline{Y}_{ia} &= \frac{1}{B} \sum_{b=1}^{B} Y_{iab}, \\ \overline{Y}_{ib} &= \frac{1}{A} \sum_{a=1}^{A} Y_{iab}, \\ \overline{Y}_{ab} &= \frac{1}{I} \sum_{i=1}^{I} Y_{iab,} \end{aligned}$$

$$\begin{aligned} \overline{Y}_A &= \frac{1}{IB} \sum_{b=1}^{B} \sum_{i=1}^{I} Y_{iab}, \\ \overline{Y}_B &= \frac{1}{IA} \sum_{a=1}^{A} \sum_{i=1}^{I} Y_{iab} \end{aligned}$$

and

$$\overline{Y}_i = \frac{1}{AB} \sum_{b=1}^{B} \sum_{a=1}^{A} Y_{iab}$$

are the sample means. It is straightforward to show that summation of squares of each side of the above equation lead to the sum of squares decomposition

$$S_t = S_a + S_b + S_i + S_{ab} + S_{ai} + S_{bi} + S_e, \tag{7.60}$$

and that terms on the right-hand side are independently distributed, where

$$S_a = IB\sum_{a=1}^{A}(\overline{Y}_a - \overline{Y})^2, \tag{7.61}$$

$$S_b = IA\sum_{b=1}^{B}(\overline{Y}_b - \overline{Y})^2, \tag{7.62}$$

$$S_i = AB\sum_{i=1}^{I}(\overline{Y}_i - \overline{Y})^2,$$

$$S_{ai} = B\sum_{i=1}^{I}\sum_{a=1}^{A}(\overline{Y}_{ia} - \overline{Y}_i - \overline{Y}_a + \overline{Y})^2,$$

$$S_{bi} = A\sum_{i=1}^{I}\sum_{b=1}^{B}(\overline{Y}_{ib} - \overline{Y}_i - \overline{Y}_b + \overline{Y})^2, \tag{7.63}$$

$$S_{ab} = I\sum_{a=1}^{A}\sum_{b=1}^{B}(\overline{Y}_{ab} - \overline{Y}_a - \overline{Y}_b + \overline{Y})^2, \tag{7.64}$$

$$S_e = \sum_{i=1}^{I}\sum_{a=1}^{A}\sum_{b=1}^{B}(Y_{iab} - \overline{Y}_{ab} - \overline{Y}_{ia} - \overline{Y}_{ib} + \overline{Y}_i + \overline{Y}_a + \overline{Y}_b - \overline{Y})^2$$

and

$$S_t = \sum_{i=1}^{I}\sum_{a=1}^{A}\sum_{b=1}^{B}(Y_{iab} - \overline{Y})^2. \tag{7.65}$$

As in the previous sections, the distribution of each sum of squares, and hence the expected value and the degrees of freedom needed to set up the ANOVA, is easily derived by averaging (7.54) over appropriate indices, writing the corresponding term in (7.54) in terms of the random effects, and then applying known results for independent terms. For example, the expected value and the distribution of S_{ab} follows from the identity

$$\overline{Y}_{ab} - \overline{Y}_a - \overline{Y}_b + \overline{Y} = \gamma_{ab} + (\overline{e}_{ab} - \overline{e}_a - \overline{e}_b + \overline{e}),$$

because from known results from the fixed effects two-way ANOVA it can be shown directly that

$$\sum_{a=1}^{A}\sum_{b=1}^{B}\frac{(\overline{e}_{ab} - \overline{e}_a - \overline{e}_b + \overline{e})^2}{\sigma^2/I} \sim \chi^2_{(A-1)(B-1)}$$

Similarly, the expected value of S_a follows from the identity

$$\overline{Y}_a - \overline{Y} = \alpha_a + \overline{\varepsilon}_a + (\overline{e}_a - \overline{e}),$$

and that of S_{ia} follows from the identity

$$\overline{Y}_{ia} - \overline{Y}_i - \overline{Y}_a + \overline{Y} = (\varepsilon_{ia} - \overline{\varepsilon}_a) + (\overline{e}_{ia} - \overline{e}_i - \overline{e}_a + \overline{e}).$$

It is now evident that the ANOVA table appropriate for model (7.54) is a mixed model version of the classical three-way ANOVA with no replicates. The degrees of freedoms, the sums of squares, and the expected values of the mean sums of squares of the resulting ANOVA are displayed in Table 7.6.

Table 7.6 RM ANOVA for the cross-classified design: Expected values

Source	DF	SS	E(MS)
Factor A	$A-1$	S_a	$\frac{IB\sum\alpha_a^2}{A-1} + \sigma_{AS}^2$
Factor B	$B-1$	S_b	$\frac{IA\sum\beta_b^2}{(B-1)} + \sigma_{BS}^2$
Subjects	$I-1$	S_i	$AB\sigma_\iota^2 + \sigma^2$
$A \times$ Subjects	$(A-1)(I-1)$	S_{ai}	$\sigma_{AS}^2 = AB\sigma_{\iota a}^2/(A-1) + \sigma^2$
$B \times$ Subjects	$(B-1)(I-1)$	S_{bi}	$\sigma_{BS}^2 = AB\sigma_{\iota b}^2/(B-1) + \sigma^2$
$A \times B$	$(A-1)(B-1)$	S_{ab}	$\frac{I\sum\sum\gamma_{ab}^2}{(A-1)(B-1)} + \sigma^2$
Error	$(A-1)(B-1)(I-1)$	S_e	σ^2
Total	$ABI-1$	S_t	

Each of the hypotheses on fixed effects, namely the equality of levels of factor A and factor B and the equality of interactions between A and B, can now be tested based on the appropriate columns of the ANOVA table. For example, the first hypothesis is tested using the F-statistic

$$F_A = \frac{\text{MSA}}{\text{MSAI}} = \frac{S_a/(A-1)}{S_{ai}/((A-1)(I-1))} \sim F_{A-1,(A-1)(I-1)}. \tag{7.66}$$

The hypothesis is rejected for large values of the F-statistic. The p-value of the test is

$$p = 1 - H_{A-1,(A-1)(I-1)}\left(\frac{S_a/(A-1)}{S_{ai}/((A-1)(I-1))}\right), \tag{7.67}$$

where $H_{(A-1),(A-1)(I-1)}$ is the cdf of the F distribution with $(A-1)$ and $(A-1)(I-1)$ degrees of freedom. As another example, the hypothesis of no interaction between A and B is tested based on the p-value

$$p = 1 - H_{(A-1)(B-1),(A-1)(B-1)(I-1)}\left(\frac{S_{ab}/((A-1)(B-1))}{S_{ai}/d_{ABI}}\right), \tag{7.68}$$

where $H_{(A-1)(B-1),d_{ABI}}$ is the cdf of the F distribution with $(A-1)(B-1)$ and

$$d_{ABI} = A-1)(B-1)(I-1)$$

degrees of freedom.

The table below summarizes the computations involved in RM ANOVA for the cross-classified design. Testing of zero variance components are also based on the F-tests provided by the ANOVA table. The software packages such as SAS, SPSS, SPlus and XPro provide procedures for performing the RM ANOVA.

RM ANOVA for the cross-classified design: F-values

Source	DF	SS	MS	F-Vale
Factor A	$A-1$	S_a	$MS_a = \frac{S_a}{A-1}$	$\frac{MS_a}{MS_{ai}}$
Factor B	$B-1$	S_b	$MS_b = \frac{S_b}{A-1}$	$\frac{MS_b}{MS_{bi}}$
Subjects	$I-1$	S_i	$MS_i = \frac{S_i}{I-1}$	$\frac{MS_i}{MS_e}$
$A \times$ Subjects	$(A-1)(I-1)$	S_{ai}	$MS_{ai} = \frac{S_{ai}}{(A-1)(I-1)}$	$\frac{MS_{ai}}{MS_e}$
$B \times$ Subjects	$(B-1)(I-1)$	S_{bi}	$MS_{bi} = \frac{S_{bi}}{(B-1)(I-1)}$	$\frac{MS_{bi}}{MS_e}$
$A \times B$	$(A-1)(B-1)$	S_{ab}	$MS_{ab} = \frac{S_{ab}}{(A-1)(B-1)}$	$\frac{MS_{ab}}{MS_e}$
Error	d_{ABI}	S_e	$MS_e = \frac{S_e}{d_{ABI}}$	
Total	$ABI-1$	S_t		

More general inferences on the variance components could also be made based on the ANOVA table. For example, the generalized p-value for testing the null hypothesis

$$H_0 : \sigma_\iota^2 \leq \sigma_0^2$$

is obtained as

$$p = 1 - EG_{I-1}\left(\frac{s_i}{AB\sigma_0^2 + s_e/V}\right), \tag{7.69}$$

where G_{N-G} is the cdf of the chi-squared distribution with $N-G$ degrees of freedom and the and expectation is taken with respect to the random variable V,

$$V = \frac{S_e}{\sigma^2} \sim \chi^2_{(A-1)(B-1)(I-1)}.$$

Generalized confidence intervals can be deduced as before using the formula for the p-value. As we discussed in Chapter 3, they could also be constructed by Tukey−Williams method with the Wang adjustment.

Example 7.3. Comparison of students' performance in Mathematics and Statistics

Table 7.7 below presents a set of test scores of 7 students taking a certain course in Mathematics and one in Statistics, which are denoted by M and S, respectively. The average score, over a scale form 0 to 10, for each course is recorded over 5 marking periods. Noting that these data comprise a cross-classified design, assume that the test scores follow model (7.54).

Table 7.7 Mathematics and Statistics scores

Course	Student	T_1	T_2	T_3	T_4	T_5
M	1	7.79	8.01	7.79	8.37	8.12
M	2	7.33	7.33	7.64	7.71	7.57
M	3	7.86	7.94	8.08	8.08	8.02
M	4	7.48	7.79	7.71	7.56	7.61
M	5	7.56	7.64	7.64	7.94	7.78
M	6	8.01	7.71	5.96	7.79	7.54
M	7	7.33	7.33	6.92	7.94	7.57
S	1	8.08	7.86	7.64	8.23	8.05
S	2	7.01	7.48	7.33	7.33	7.30
S	3	8.08	8.08	8.08	8.37	8.23
S	4	7.71	7.79	7.79	8.01	7.89
S	5	7.64	7.86	7.79	7.48	7.63
S	6	6.50	6.41	5.96	6.24	6.27
S	7	6.92	6.92	7.71	7.71	7.46

Data can be conveniently analyzed using a software package that provides repeated measures procedures. For instance, SAS PROC GLM with the REPEATED statement or XPro can be employed to obtain the ANOVA table. The F-values and p-values for testing fixed effects, computed using (7.66) and (7.68) are shown in the following ANOVA table. It also shows the F-values and p-values for testing the variance components. It should be emphasized that the latter p-values are appropriate for testing the null hypotheses that the variance components are equal to zero only.

Source	DF	SS	F-value	p-value
Course	1	0.500	0.884	0.383
Period	4	0.909	2.151	0.105
Students	6	10.80	29.80	0.000
Course × Students	6	3.051	8.416	0.000
Period × Students	24	2.534	1.747	0.089
Course × Period	4	0.274	1.760	0.364
Error	24	1.450		
Total	68	19.467		

Obviously, according to the observed p-values, 0.383, 0.105, and 0.36, none of the fixed effects, the mean scores for courses, the effects of marking periods, or their interactions are significantly different. The variance components due to among-subject variation and their interaction with the subjects are highly significant. The interaction between the subjects and the periods, however, is not quite significant at the 0.05 level. The p-value suggests some week significance of that variance component as well. The following table shows the 95% generalized confidence intervals and the adjusted Tukey−Williams intervals, in terms of lower bounds (LB) and upper bounds (UB) of the intervals, for the three variance components. The confidence intervals also imply the above conclusions concerning variance components. Notice also that the adjusted Tukey−Williams intervals and the generalized intervals are practically the same.

	Adjusted T−W Interval		**Generalized Interval**	
Variance Comp.	LB	UB	LB	UB
Students	0.067	0.867	0.068	0.866
Course × Students	0.014	0.241	0.014	0.240
Period × Students	0.0	0.076	0.0	0.072

7.5.2 Nested design

Now suppose the levels of factor B are nested under the levels of factor A. This is the case, for instance, when an experiment is carried out in a number of phases and levels of factor B are the occasions at which measurements are taken. In each phase of the experiment, the level of Factor A, say the treatment is kept fixed at a given level and observations are taken from each of the I subjects. For the sake of the simplicity of notation, let B denote the total number of levels of factor B, so that the total number of observations is BI. When we have a set of data according to a nested design of this nature, the data can be set out as in Table 7.8.

Table 7.8 Layout of nested design

	A_1			$\cdots$		A_a		$\cdots$		A_A	
Subject	B_1	$\cdots$	$B_.$	$\cdots$	$B_.$	$\cdots$	$B_.$	$\cdots$	$B_.$	$\cdots$	B_B
1	$Y_{\cdots}$	$\cdots$	$Y_{\cdots}$	$\cdots$	$Y_{\cdots}$	$\cdots$	$Y_{\cdots}$	$\cdots$	$Y_{\cdots}$	$\cdots$	$Y_{\cdots}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	$Y_{\cdots}$	$\cdots$	$Y_{\cdots}$	$\cdots$	$Y_{\cdots}$	$\cdots$	$Y_{\cdots}$	$\cdots$	$Y_{\cdots}$	$\cdots$	$Y_{\cdots}$

Since the levels of factor B are nested under that of A, here we can allow only main effects of A and the interaction between A and B. Similarly, without causing identification problems, in addition to the error term we can introduce up to two random effects in a standard mixed model. So, a model appropriate for this design is

$$Y_{ib(a)} = \mu + \alpha_a + \beta_{b(a)} + \epsilon_i + \varepsilon_{ia} + e_{ib(a)}, \tag{7.70}$$

$a = 1, \ldots, A; b = 1, \ldots, b_a;\ i = 1, \ldots, I$, where α_a and $\beta_{b(a)}$ are fixed effects, ϵ_i is the random effect due to subjects, ε_{ia} is the random effect representing differential subject random effects at different levels of factor A, $e_{ib(a)}$ is the residual error, and b_a is the number of observations available from each subject when factor A takes on the value a. As before we assume that

$$\epsilon_i \sim N(0, \sigma_\iota^2), \varepsilon_{ia} \sim N(0, \sigma_{\iota a}^2) \quad \text{and} \quad e_{iab} \sim N(0, \sigma^2). \tag{7.71}$$

and that they are independently distributed. Assume that the parameters are normalized so that, when summed over the levels of fixed effects, they sum to zero for all summations needed in the decomposition of the sums of squares. Moreover, ε_{ia} being an incremental random effect we assume that $\sum_{a=1}^{A} \varepsilon_{ia} = 0$. As before, define the sample means in the obvious manner as

$$\begin{aligned} \overline{Y}_{ia} &= \frac{1}{b_a} \sum_{b(a)=1}^{b_a} Y_{ib(a)}, \\ \overline{Y}_{b(a)} &= \frac{1}{I} \sum_{i=1}^{I} Y_{ib(a)}, \end{aligned}$$

$$\begin{aligned} \overline{Y}_i &= \frac{1}{A} \sum_{a=1}^{A} \overline{Y}_{ia}, \\ \overline{Y}_a &= \frac{1}{I} \sum_{i=1}^{I} \overline{Y}_{ia}, \end{aligned}$$

and

$$\overline{Y} = \frac{1}{A} \sum_{a=1}^{A} \overline{Y}_a.$$

Define the sums of squares as

$$S_i = B \sum_{i=1}^{I} (\overline{Y}_i - \overline{Y})^2, \tag{7.72}$$

$$S_a = I \sum_{a=1}^{A} b_a (\overline{Y}_a - \overline{Y})^2, \tag{7.73}$$

$$S_{ia} = \sum_{i=1}^{I}\sum_{a=1}^{A} b_a(\overline{Y}_{ia} - \overline{Y}_i - \overline{Y}_a + \overline{Y})^2, \tag{7.74}$$

$$S_{b(a)} = I\sum_{a=1}^{A}\sum_{b(a)=1}^{b_a}(\overline{Y}_{b(a)} - \overline{Y}_a)^2, \tag{7.75}$$

$$S_e = \sum_{i=1}^{I}\sum_{a=1}^{A}\sum_{b(a)=1}^{b_a}(Y_{ib(a)} - \overline{Y}_{ia} - \overline{Y}_{b(a)} + \overline{Y}_a)^2 \tag{7.76}$$

and

$$S_t = \sum_{i=1}^{I}\sum_{a=1}^{A}\sum_{b(a)=1}^{b_a}(Y_{ib(a)} - \overline{Y})^2. \tag{7.77}$$

The ANOVA for model (7.70) can be constructed using these sums of squares.

The distribution of the sums of squares can be derived by expressing them in terms of parameters and independent random variables appearing on the right-hand side of (7.70). For example, they follow from the identities

$$\overline{Y}_{b(a)} - \overline{Y}_a = \beta_{b(a)} + (\overline{e}_{b(a)} - \overline{e}_a),$$

$$\overline{Y}_{ia} - \overline{Y}_i - \overline{Y}_a + \overline{Y} = (\varepsilon_{ia} - \overline{\varepsilon}_a) + (\overline{e}_{ia} - \overline{e}_i - \overline{e}_a + \overline{e}),$$

$$Y_{ib(a)} - \overline{Y}_{ia} - \overline{Y}_{b(a)} + \overline{Y}_a = (e_{ib(a)} - \overline{e}_{ia} - \overline{e}_{b(a)} + \overline{e}_a),$$

$$\overline{Y}_i - \overline{Y} = (\epsilon_i - \overline{\epsilon}) + (\overline{e}_i - \overline{e})$$

and

$$\overline{Y}_a - \overline{Y} = \alpha_a + \overline{\varepsilon}_a + (\overline{e}_a - \overline{e})$$

Actually, the current design can be thought of as a variation of the one considered in Section 7.3 and deduce the ANOVA table. In fact, the degrees of freedom and the sums of squares necessary to construct the ANOVA table for the current design could be obtained from any software package, which provides the ANOVA for the model considered in that section. Since the model is different, however, the definitions of sources of variation, the interpretations, and the tests based on the ANOVA table are different. Here it is the measurement occasions or the levels of factor B that are nested under treatment groups (levels of factor A) rather than the subjects. Table 7.9 provides the ANOVA for making usual inferences about the model (7.70).

The procedure of testing the fixed effects and the variance components based on the ANOVA table is similar to that of the previous section. For example, the equality of the effects of A levels is tested using the F-statistic

$$F_1 = \frac{S_a/(A-1)}{S_{ia}/((I-1)(A-1))} \sim F_{A-1,(I-1)(A-1)} \tag{7.78}$$

Table 7.9 RM ANOVA for the nested design: Expected values

Source	DF	SS	E(MS)
Subjects	$I-1$	S_i	$B\sigma_i^2+\sigma^2$
Intereaction $B(A)$	$B-A$	$S_{b(a)}$	$I\sum_{a=1}^{A}\sum_{b\in a}\beta_{b(a)}^2/(B-A)+\sigma^2$
Factor A	$A-1$	S_a	$I\sum_{a=1}^{A}\sum_{b\in a}\alpha_a^2/(A-1)+\sigma_{AW}^2$
$A\times$ Subjects	$(I-1)(A-1)$	S_{ia}	$\sigma_{AW}^2=B\sigma_{\iota a}^2/(A-1)+\sigma^2$
Occasions $\times$ Subjects	$(I-1)(B-A)$	S_e	σ^2
Total	$BI-1$	S_t	

and the equality of $\beta_{b(a)}$ terms are tested using the $F-$statistic

$$F_2=\frac{S_{b(a)}/(B-A)}{S_e/((I-1)(B-A))}\sim F_{B-A,(I-1)(B-A)}\ . \tag{7.79}$$

The table below summarizes the computations involved in RM ANOVA for the nested design.

RM ANOVA for the nested design: F-values

Source	DF	SS	MS	F-Value
Subjects	$I-1$	S_i	$MS_i=\frac{S_i}{I-1}$	$\frac{MS_i}{MS_e}$
Interaction $B(A)$	$B-A$	$S_{b(a)}$	$MS_{b(a)}=\frac{S_{b(a)}}{B-A}$	$\frac{MSb(a)}{MS_e}$
Factor A	$A-1$	S_a	$MS_a=\frac{S_a}{A-1}$	$\frac{MS_a}{MS_{ia}}$
$A\times$ Subjects	$(I-1)(A-1)$	S_{ia}	$MS_{ia}=\frac{S_{ia}}{(I-1)(A-1)}$	$\frac{MS_{ia}}{MS_e}$
Occa. $\times$ Sub.	$(I-1)(B-A)$	S_e	$MS_e=\frac{S_e}{(I-1)(B-A)}$	
Total	$BI-1$	S_t		

The testing of zero variance components is also based on F-tests provided by the above ANOVA. Other inferences are carried out by the generalized approach based on the distributional results

$$\frac{S_e}{\sigma^2}\sim\chi^2_{(I-1)(B-a)},$$

$$\frac{S_i}{B\sigma_i^2+\sigma^2}\sim\chi^2_{(I-1)}\ ,$$

and

$$\frac{S_{ia}}{B\sigma_{\iota a}^2/(A-1)+\sigma^2}\sim\chi^2_{(I-1)(A-1)}\ .$$

7.6 REGRESSION AND RM ANCOVA

In Section 7.3 we already saw the need for regression methods to reduce the among subjects variation. We also need to take the regression approach when some of the important factors affecting the response variable do not take on a set of discrete or qualitative values, as was the case in the foregoing treatment. In a regression setting, we can handle both situations by introducing appropriately defined covariates in the kind of models we discussed above. We will continue to assume that the covariates have an intrinsically linear (which includes polynomials defined in terms of the covariates) relationship with the response variable.

7.6.1 Controlling subject variation

First consider the situation where the objective of the regression is to control some of the variation of the experimental units. Invoking regression methods to reduce the among-subject variation is especially important when the response variable substantially vary from one subject to another. While biomedical experiments are not really exceptions, this problem arise more often in socioeconomic, marketing, environmental, and agricultural experiments. For example, when a market analyst needs to test the efficacy and compare alternative promotions on a product, the analyst would obtain a set of longitudinal data from a sample of stores selling the product for a number of time periods before and after the promotions. The analyst would also use some control stores with no promotion. In this application there could be a substantial variation among store sales. This is because there could be large and small stores and various types of stores (e.g. supermarkets and drug stores) selling the product. Then, in analyzing the data from the marketing experiment, the analyst needs to tackle the problem of store heterogeneity. In fact the problem is likely to arise in any situation where the experimental unit is a store, a school, a hospital, a plot of land, a geographical region such as a city, and so on.

The easiest and perhaps one of the most effective way of controlling the among subject variation is using a set of historical data before the treatments are administered. When historical data are not available one can also use certain attributes of the experimental unit to control the variation. One can do this even when historical data are available and further controlling is desirable. For example, in the marketing experiment described above, the number of checkouts of a store can be used to represent the store size. This store attribute and other attributes such as the type of a store (supermarket, drug store, convenience store, etc.) are also important in formulating a regression model when we are interested in estimating the effect of the promotion by store attributes. It is indeed the case that the effect of a promotion at a supermarket is different from that at a convenience store. In a regression setting we can easily handle both the quantitative variables and the qualitative variables together, when the latter is represented by a set of dummy variables.

In the treatment below we assume that the data have been transformed so that we can assume a simple linear regression for modeling data from a group of subjects. Koschat and Weerahandi (2003) showed that widely used log transformed data are not good enough to achieve the homoscedastic variances required in the ordinary least-squares estimation of regression parameters. Figure 7.2 below shows a plot of a representative sales data S from the above application transformed as $Y_{i(g)} = \log(S_{i(g)}) - \overline{lS}$, where $\overline{lS}$ is the sample mean of $\log(S_{i(g)})$ as a function of the store size. Observe that the variance of data tend to decrease dramatically with the size of the stores. Ignoring this fact will result in highly unreliable results due to large variation of small stores. In other words, ordinary least-squares estimation of parameters would give equal weights to the observations from small and large stores, whereas Figure 7.2 suggests that we should give much smaller weights to the smaller stores. To outline some useful results dealing with heterogeneous experimental units, consider the problem of estimating a vector of parameters $\boldsymbol{\lambda}$ based on a vector of response variables $\mathbf{S}$ and a matrix of covariates $\mathbf{W}$. Koschat and Weerahandi (2003) showed that when a set of historical values on $\mathbf{S}$ is available, a model appropriate to handle the data can be set up as $\mathbf{S} = \mathbf{U}\boldsymbol{\lambda}+\mathbf{E}$, where $\mathbf{U}_{i(g)}=\overline{S}_{i(g)}W_{i(g)}$ and $\overline{\mathbf{S}}$ is the vector of mean historical responses. They showed that this particular transformation leads to an error structure with known weights that allows us to transform the model to a regression model with a homoscedastic error structure. More specifically, they showed the transformation leads to an error structure of the form $\mathbf{E}\tilde{\ }N\left(\mathbf{0},\sigma^2\mathbf{D}\right)$, where $\mathbf{D} = diag(\overline{S}_1,\cdots,\overline{S}_n)$, which leads to the simple linear regression model

$$\widetilde{\mathbf{S}} = \widetilde{\mathbf{U}}\boldsymbol{\lambda} + \widetilde{\mathbf{E}}, \text{ where } \widetilde{\mathbf{E}} = \mathbf{D}^{-1/2}\mathbf{E}\tilde{\ }N\left(\mathbf{0},\sigma^2\mathbf{I}\right)$$

and $\widetilde{\mathbf{S}} = \mathbf{D}^{-1/2}\mathbf{S}$, $\widetilde{\mathbf{U}} = \mathbf{D}^{-1/2}\mathbf{U}$, where

$$\mathbf{D}^{-1/2} = \begin{pmatrix} \overline{S}_1^{-1/2} & 0 & \vdots & 0 \\ 0 & \overline{S}_2^{-1/2} & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & \overline{S}_n^{-1/2} \end{pmatrix}$$

is the inverse square root of matrix $\mathbf{D}$.

7.6.2 Classical ANCOVA

To introduce the method of Analysis of Covariance (ANCOVA), first consider the problem of testing the overall effect of the treatment during the trial period, possibly by one or more other factors, when the dependent variable of the

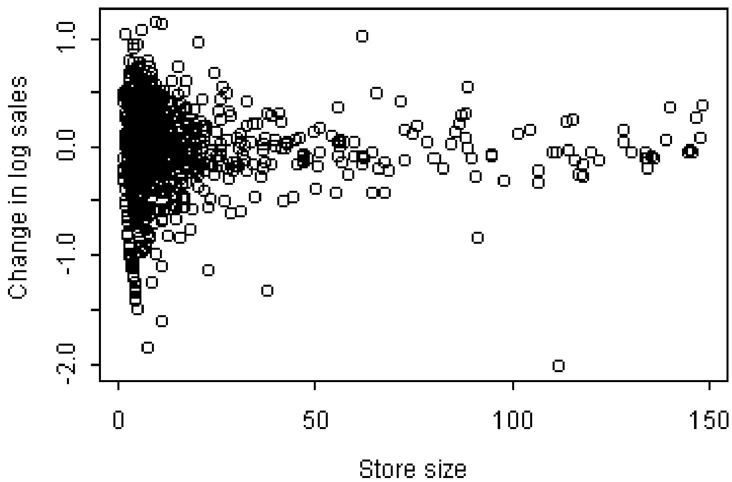

Figure 7.2 Heteroscedasticity of log sales data

regression is the average of responses taken over time. The results presented here also applies when we need to estimate and compare the treatment effects for each time period separately, without MANOVA and without making any assumption on the correlation structure for the responses taken over time. In this case, multiple comparisons are handled with the Bonferroni adjustment. Let $Y_{i(g)}$ be the response from ith subject in group g and $X_{ki(g)}$ be the value of kth covariate to be used in the regression. One of the covariates is possibly the average of historical data obtained from the subject prior to administering of the treatments. To compare G groups, consider the linear model

$$Y_{i(g)} = \theta_g + \sum_{k=1}^{K} \beta_k X_{ki(g)} + \epsilon_{i(g)}; \tag{7.80}$$

$$i(g) = 1, \ldots,\ n_g; g = 1, \ldots, G.$$

As before assume that

$$\epsilon_{i(g)} \sim N(0, \sigma^2), \tag{7.81}$$

an assumption that can be relaxed by invoking the methods suggested by Koschat and Weerahandi (1990, 1992) and Ananda (1998). Let

$$\mathbf{y}'_g = \left(Y_{g1}, Y_{g2}, ..., Y_{gn_g}\right)$$

be the responses available from Group g and let $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, ..., \mathbf{y}'_G)'$ be the vector of all $N = \sum n_g$ responses. Similarly, let $\mathbf{X}_g$ denote the $n_{g.} \times K$ matrix of covariate data for the subjects in Group g, and $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, ..., \mathbf{X}'_G)'$ is a matrix of dimension $N_. \times K$. Then, in matrix notation, the model can be expressed as

$$\begin{aligned} \mathbf{y} &= \mathbf{Z}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right) \\ &= \mathbf{W}\boldsymbol{\lambda} + \boldsymbol{\epsilon} \end{aligned} \tag{7.82}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_G)'$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_K)'$, $\boldsymbol{\lambda} = (\boldsymbol{\theta}, \boldsymbol{\beta})'$,

$$\mathbf{Z} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1}_{n_G} \end{pmatrix}_{N \times G}$$

is the design matrix formed by the dummy variables indicating the groups, and

$$\mathbf{W} = (\mathbf{Z}, \mathbf{X}).$$

Moreover, we assume that the data have already been transformed appropriately before setting up model (7.82). As discussed by Koschat and Weerahandi (2003) this is especially important in dealing with data from heterogeneous experimental units such as stores, cities, schools, hospitals, and so on.

Consider the problem of testing the hypothesis

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_G. \tag{7.83}$$

If the null hypothesis is true, then the model reduces to

$$\begin{aligned} \mathbf{y} &= \theta + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &= \mathbf{V}\widetilde{\boldsymbol{\beta}} + \boldsymbol{\epsilon}, \end{aligned} \tag{7.84}$$

where θ is the common parameter under the null hypothesis, $\mathbf{V} = (\mathbf{1}_N, \mathbf{X})$, and $\widetilde{\boldsymbol{\beta}} = (\theta, \boldsymbol{\beta}')'$. The Analysis of Covariance (ANCOVA) is based on the reduction of error sum of squares from model (7.84) to model (7.82). The error sum of squares for model (7.82) is easily computed from $\mathbf{e} = \mathbf{y} - \mathbf{W}\widehat{\boldsymbol{\lambda}}$ or from the standard regression formula as

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= \mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\lambda}}'\mathbf{W}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y} \\ &= \mathbf{y}'(\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{y}. \end{aligned}$$

The error sum of squares under (7.84) is computed by replacing $\mathbf{W}$ in above formula by $\mathbf{V}$. The degrees of freedom of these quantities can also be deduced from the standard regression results. The results lead to an F-test and all necessary computations can be summarized in an ANCOVA table as illustrated by Table 7.10, where the sums of squares are defined as

$$\begin{aligned} S_1 &= \mathbf{y}'(\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' - \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}')\mathbf{y}. \\ S_2 &= \mathbf{y}'(\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-\mathbf{1}}\mathbf{W}')\mathbf{y}, \end{aligned}$$

and

$$S_3 = \mathbf{y}'(\mathbf{I} - \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}')\mathbf{y}.$$

.

Table 7.10 ANCOVA for comparing treatment groups

Source	DF	SS	MS	F-Value
Groups	$G-1$	S_1	$MS_1 = \frac{S_i}{G-1}$	$\frac{MS_1}{MS_2}$
Error with $\mathbf{W} = (\mathbf{Z}, \mathbf{X})$	$N-G-K$	S_2	$MS_2 = \frac{S_2}{N-G-K}$	
Error with $\mathbf{V} = (\mathbf{1}_N, \mathbf{X})$	$N-K-1$	S_3		

It is evident from the ANCOVA table that the p-value for testing H_0 is

$$p = 1 - H_{G-1,N-G-K}\left(\frac{N-G-K}{G-1}\frac{\mathbf{y}'(\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' - \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}')\mathbf{y}}{\mathbf{y}'(\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-\mathbf{1}}\mathbf{W}')\mathbf{y}}\right), \tag{7.85}$$

where $H_{G-1,N-G-K}$ is the cdf of the F distribution with $G-1$ and $N-G-K$ degrees of freedom. The null hypothesis of equal treatment effects is rejected for small values of the p-value.

Example 7.4. Testing the efficacy of promotions

Table 7.11 below shows a set of illustrative data on sales of a cellular service plan at 15 stores during a promotional campaign period of 5 weeks. At five of the test stores a free phone is given as a purchase incentive and at five other stores, the service activation fee is waived during the trial period. The rest of the stores are kept as control stores with no special promotion. The three groups of stores are denoted as P, A, and C. Let us use the same notation to indicate the 1-0 dummy variables representing the groups. Although the data in the table are hypothetical, they are generated from a model of the form assumed above. Also shown in the table are mean sales denoted by $\overline{\mathbf{Y}}_{post}$ at each store during the trial period. In this type of application it is important to control the store to store variation using a set of historical sales. In Table 7.9 a set of historical sales from the stores are also given and are denoted by $\overline{\mathbf{Y}}_{pre}$.

Let us first consider the problem of comparing the overall effects of the promotion using a regression of $\overline{\mathbf{Y}}_{post}$ on $\overline{\mathbf{Y}}_{pre}$ and group dummies. More precisely, consider the regression model

$$\overline{\mathbf{Y}}_{post} = \alpha\overline{\mathbf{Y}}_{pre} + \beta C + \gamma A + \lambda P + \epsilon.$$

The table below shows the estimated parameters of the model along with the standard errors of the estimates and the corresponding t-values. The

Table 7.11 Cellular service sales by store and week

	Time:	*pre*	t_1	t_2	t_3	t_4	t_5	*post*
Group	**Store**	$\overline{\mathbf{Y}}_{pre}$	$\mathbf{Y}_1$	$\mathbf{Y}_2$	$\mathbf{Y}_3$	$\mathbf{Y}_4$	$\mathbf{Y}_5$	$\overline{\mathbf{Y}}_{post}$
C	1	406.5	423	406	412	407	426	414.8
C	2	154.3	183	185	155	177	169	173.8
C	3	491.8	499	507	454	507	495	492.4
C	4	61.6	65	82	88	65	80	76
C	5	384.8	394	375	403	378	374	384.8
A	6	603.3	630	625	620	620	693	637.6
A	7	74.6	19	175	118	61	197	114
A	8	403.9	316	457	482	487	444	437.2
A	9	68.4	93	57	186	79	116	106.2
A	10	487.5	495	502	554	570	503	524.8
P	11	294.4	357	417	416	527	318	407
P	12	173.4	245	239	226	266	372	269.6
P	13	580.7	514	615	760	715	765	673.8
P	14	113.2	101	235	203	237	332	221.6
P	15	172.2	374	351	253	384	245	321.4

table suggests that the promotions has worked. This can be formally tested performing an ANCOVA.

Variable	Parameter	Estimate	Std. Error	t-value
$\overline{\mathbf{Y}}_{pre}$	α	0.9718	0.0184	52.7704
C	β	17.0220	8.0772	2.1074
A	γ	45.6649	8.4347	5.4140
P	λ	119.4300	7.6745	15.5620

This can be accomplished by running the regression $\overline{\mathbf{Y}}_{post} = \delta + \alpha\overline{\mathbf{Y}}_{pre} + \epsilon$ in the absence of the effects due to promotions and studying the increase in error sum of squares. The ANCOVA table obtained using the results of the two regressions is shown below.

ANCOVA for testing the promotional effects

Source	DF	SS	MS	F-Value
Groups	2	25159.9	12579.95	73.27
Error with $(\mathbf{C}, \mathbf{A}, \mathbf{P}, \overline{\mathbf{Y}}_{pre})$	11	1912.0	173.81	
Error with $(\mathbf{1}_{15}, \overline{\mathbf{Y}}_{pre})$	13	27071.9		

Based on the results in the ANCOVA table the p-value for testing the hypothesis of no effect due to promotions can be computed as $p = 1 - H_{2,11}(73.27) = 0.0$. According to the p-value and the estimated parameters of the regression model, the promotions are highly significant. Moreover, we can now proceed to compare the two promotions to conclude by means of the t-test that the free phone has been more effective than the waiver of the service activation fee. If the cost of each incentive is available, one can also investigate the economics of the two incentive plans.

7.6.3 RM ANCOVA

In the previous section we assumed that we have only one observation taken from each subject. Now suppose we have repeated measures taken over time from each subject as was the case in Section 7.3. Also assume that the matrix $\mathbf{X}$ of covariates is measured only once as the case in many applications including the problem of handling that subject heterogeniety with historical mean responses that motivated the treatment in this section. Also we continue to assume that the regression coefficients are the same for all groups. The readers interested in results for situations when these assumptions do not hold are referred to Vonesh and Chinchilli (1997).

By combining models(7.82) and (7.80), we then get

$$\begin{aligned} Y_{i(g)t} &= \theta_g + \lambda_t + \gamma_{gt} + \sum_{k=1}^{K} \beta_k X_{ki(g)} + \varepsilon_{i(g)t} \qquad (7.86) \\ t &= 1, \ldots, T;\ i(g) = 1, \ldots,\ n_g; g = 1, \ldots, G, \end{aligned}$$

where θ_g is the mean effect of treatment g, λ_t is the occasion effect at time t, γ_{gt} is their interaction, $\varepsilon_{i(g)t} = \alpha_{i(g)} + \epsilon_{i(g)t}$, and $\varepsilon_{i(g)} = (\ \varepsilon_{i(g)1}, \ldots,\ \varepsilon_{i(g)T})'$ is distributed as $\varepsilon_{i(g)} \sim N(0, \Sigma)$. As seen in Section 7.3, the error covariance matrix Σ has a compound symmetric covariance structure of the form

$$\Sigma = \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T^{'} + \sigma^2 \mathbf{I}_T.$$

A major advantage of a covariance matrix having the compound symmetric structure is that the generalized least-squares estimates (also the MLEs) of parameters are the same as the ordinary least-squares estimates. Also assume that λ_t and γ_{gt} satisfy the constraints given in Section 7.3. Then, as in that section, inferences on treatment effects and the coefficients of covariates can be based on the data

$$\begin{aligned} \overline{Y}_{i(g)} &= \theta_g + \sum_{k=1}^{K} \beta_k X_{ki(g)} + \overline{\varepsilon}_{i(g)} \qquad (7.87) \\ &= \theta_g + \mathbf{X}'_{i(g)} \boldsymbol{\beta} + \overline{\varepsilon}_{i(g)}, \qquad (7.88) \end{aligned}$$

a result that can also be established by a transformation of the data with an orthogonal matrix having the vector $T^{-1/2}\mathbf{1}'_T$ as its first row, where $\mathbf{X}'_{i(g)} = (X_{1i(g)}, X_{2i(g)}, ..., X_{Ki(g)})$ and

$$\overline{Y}_{i(g)} = \frac{1}{T}\sum_{t=1}^{T} Y_{i(g)t} \ .$$

As before, the model can be expressed in matrix notation as

$$\begin{aligned} \overline{\mathbf{y}} &= \mathbf{Z}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \widetilde{\sigma}^2\mathbf{I}\right) \\ &= \mathbf{W}\boldsymbol{\lambda} + \boldsymbol{\epsilon}, \end{aligned} \tag{7.89}$$

where $\overline{\mathbf{y}} = (\overline{\mathbf{y}}'_1, \overline{\mathbf{y}}'_2, ..., \overline{\mathbf{y}}'_G)'$ and $\overline{\mathbf{y}}'_g = \left(\overline{Y}_{g1}, \overline{Y}_{g2}, ..., \overline{Y}_{gn_g}\right)$. Model (7.89) allows us to compare the treatments as in the conventional ANCOVA. To test the significance of occasion effects and interactions, we obtain from (7.86) and (7.87)

$$Y_{i(g)t} - \overline{Y}_{i(g)} = \lambda_t + \gamma_{gt} + \varepsilon'_{i(g)t}.$$

Now it is clear that tests on the significance of occasion effects and interactions can be based on the sum of squares decomposition implied by the identity

$$\begin{aligned} (Y_{i(g)t} - \overline{Y}_{i(g)}) &= (\overline{Y}_t - \overline{Y}) + (\overline{Y}_{gt} - \overline{Y}_t - \overline{Y}_g + \overline{Y}) \\ &\quad + (Y_{i(g)t} - \overline{Y}_{gt} - \overline{Y}_{i(g)} + \overline{Y}_g), \end{aligned} \tag{7.90}$$

where various sample means are defined as in Section 7.3. To construct the RM ANCOVA in view of these results, define

$$\begin{aligned} S_1 &= \overline{\mathbf{y}}'(\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' - \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}')\overline{\mathbf{y}}. \\ S_2 &= \overline{\mathbf{y}}'(\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\overline{\mathbf{y}}, \end{aligned}$$

$$S_o = N\sum_{t=1}^{T}(\overline{Y}_t - \overline{Y})^2,$$

$$S_{og} = \sum_{t=1}^{T}\sum_{g=1}^{G} n_g(\overline{Y}_{gt} - \overline{Y}_t - \overline{Y}_g + \overline{Y})^2,$$

and

$$S_e = \sum_{t=1}^{T}\sum_{g=1}^{G}\sum_{i\in g}(Y_{i(g)t} - \overline{Y}_{gt} - \overline{Y}_{i(g)} + \overline{Y}_g)^2.$$

Note that the definitions of the sums of squares needed in testing occasion effects and interactions are the same as those used in RM ANOVA. With this notation, we can now carry out the RM ANCOVA as summarized by Table 7.12.

RM ANCOVA for testing group and occasion effects

Source	DF	SS	MS	F-Val
Groups	$G-1$	S_1	$MS_1 = \frac{S_i}{G-1}$	$\frac{MS_1}{MS_2}$
Error with $\mathbf{W}$	$N-G-K$	S_2	$MS_2 = \frac{S_2}{N-G-K}$	
Occasions	$T-1$	S_o	$MS_o = \frac{S_o}{T-1}$	$\frac{MS_o}{MS_e}$
Groups × Occa.	$(T-1)(G-1)$	S_{og}	$MS_{og} = \frac{S_{og}}{(T-1)(G-1)}$	$\frac{MS_{og}}{MS_e}$
Error	$(T-1)(N-G)$	S_e	$MS_e = \frac{S_e}{(T-1)(N-G)}$	

Example 7.5. Testing the efficacy of promotions (continued)

Consider again the data in Table 7.11 representing the sales of a cellular service plan at 15 stores during a promotional campaign. We are now in a position to further study the effects of the promotions. Since $\overline{\mathbf{Y}}_{pre}$ is the only covariate beyond the group effects used in the regression of Example 7.4, according to (7.89), the F-test of the hypothesis of no effect due to promotions is the same as the one reported in Example 7.4. According to (7.90), the rest of the information needed to set up the RM ANCOVA table can be obtained by performing a classical RM ANOVA on the data. The complete RM ANCOVA obtained in this manner is shown below.

Table 7.12 RM ANCOVA for testing the effects due to week and promotion

Source	DF	SS	MS	F-Value
Groups	2	25159.9	12579.95	73.27
Error with $(\mathbf{C}, \mathbf{A}, \mathbf{P}, \overline{\mathbf{Y}}_{pre})$	11	1912.0	173.81	
Occations	4	28750.9	7187.73	2.47
Groups × Occations	8	27042.0	3380.25	1.17
Error	48	139719.0	2910.81	

As we had concluded in Example 7.5, the differences in group effects are highly significant and that the promotions have worked in increasing the sales of cellular service. The p-value based on the F-value of the RM ANCOVA table for testing the differences in occasion effects is 0.057, which suggests that there is some evidence of changed (in fact increased) sales over time. The p-value for testing the hypothesis of no interaction effect is 0.3417. These p-values lead us to conclude that, while the promotions have worked and have been steady over time, the sales have been somewhat increasing over the trail period and that the trends are about the same for all groups.

Exercises

7.1 Consider the repeated measures problem with one group of subjects

$$Y_{it} = \alpha_i + \beta_t + \epsilon_{it} \ ,$$

where α_i is the random effect due to subject i, β_t, $t = 1, \ldots, T$ are fixed effects due to occasions/treatments, and ϵ_{it} are the residual terms. Also make the usual normality assumption that

$$\alpha_i \sim N(0, \sigma_\alpha^2), \quad \epsilon_{it} \sim N(0, \sigma^2); \quad t = 1, \ldots, T; \ i = 1, \ldots, I$$

and that they are all independently distributed.

(a) Show hat $\mathrm{Var}(Y_{it}) = \sigma_\alpha^2 + \sigma^2$ and that $\mathrm{Cov}(Y_{it}, Y_{it'}) = \sigma_\alpha^2$, and hence that the data vector formed by observations taken from each subject has a covariance matrix with the compound symmetric structure.
(b) Consider the decomposition of sample means,

$$Y_{it} - \overline{Y} = (Y_i - \overline{Y}) + (Y_t - \overline{Y}) + (Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y}).$$

(c) Show that

$$Y_t - \overline{Y} = (\beta_t - \overline{\beta}) + (\epsilon_t - \overline{\epsilon})$$

and hence obtain distribution of the mean sum of squares, $MS_B = I \sum (Y_t - \overline{Y})^2 \ /(T-1)$ and deduce that it has mean $I \sum (\beta_t - \overline{\beta})^2 /(T-1) + \sigma^2$.
(d) Similarly express each other term in the decomposition in terms of parameters and random variables with zero means, and derive the distribution of the corresponding mean sum of squares.

7.2 Consider again the model in Exercise 7.1. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)'$ be the vector formed by all occasion means and let $\mathbf{a}$ be a vector of known constants of same dimension. Establish procedures for testing hypotheses of the form

$$H_0 : \mathbf{a}'\boldsymbol{\beta} \leq k,$$

where k is a hypothesized constant. Deduce procedures for constructing $100\gamma\%$ confidence intervals for equal-tail confidence intervals for $\mathbf{a}'\boldsymbol{\beta}$.

7.3 Consider the data set reported in Table 7.1. Assuming model 7.1, test the hypothesis that the mean reaction times for probe words 1 is 8 units greater than that for probe words 2. Construct a 99% confidence interval for the mean difference. Also construct 99% confidence intervals for the sum of all occasion means and for each variance component.

7.4 Consider model (7.20) and the decomposition

$$\begin{aligned}(Y_{i(g)t} - \overline{Y}) \quad &= \quad (\overline{Y}_{i(g)} - \overline{Y}) + (\overline{Y}_t - \overline{Y}) \\ &\quad + (\overline{Y}_{gt} - \overline{Y}_t - \overline{Y}_g + \overline{Y}) + (Y_{i(g)t} - \overline{Y}_{gt} - \overline{Y}_{i(g)} + \overline{Y}_g).\end{aligned}$$

(a) Show that

$$Y_{i(g)t} - \overline{Y}_{gt} - \overline{Y}_{i(g)} + \overline{Y}_g = e_{i(g)t} - \overline{e}_{gt} - \overline{e}_{i(g)} + \overline{e}_g.$$

(b) Hence show that

$$\frac{\sum_{t=1}^{T}\sum_{g=1}^{G}\sum_{i\in g}(Y_{i(g)t} - \overline{Y}_{gt} - \overline{Y}_{i(g)} + \overline{Y}_g)^2}{\sigma^2} \sim \chi^2_{(N-G)(T-1)}.$$

(c) Similarly express each of the other terms in the decomposition in terms of parameters and random variables with zero means, and derive the expected value and the distribution of the corresponding mean sum of squares.

7.5 Consider the data in Table 7.4 and assume model (7.20).

(a) Construct 95% equal-tail confidence interval for the mean difference between the first two treatment groups.
(b) Construct 99% equal-tail confidence intervals for the variance component.
(c) Test the hypothesis that the variance component is greater than 600.

7.6 Consider model (7.54). By taking an approach similar to that of Section 7.3 or otherwise, establish procedures for multiple comparisons.

7.7 Consider the data in Table 7.7 and assume model (7.54).

(a) Test the hypothesis that subject's variance component is less than 1.
(b) Construct a 95% equal-tail confidence interval for the difference in mean scores in Statistics and Mathematics.
(c) Construct 99% equal-tail confidence intervals for each variance component.
(d) Construct 95% lower confidence bounds for each of the variance components.

7.8 With the usual notation, consider the model,

$$Y_{ib(a)} = \mu + \alpha_a + \beta_{b(a)} + \epsilon_i + \varepsilon_{ia} + e_{ib(a)}.$$

(a) Show that

$$\overline{Y}_{ia} - \overline{Y}_i - \overline{Y}_a + \overline{Y} = (\varepsilon_{ia} - \overline{\varepsilon}_a) + (\overline{e}_{ia} - \overline{e}_i - \overline{e}_a + \overline{e}).$$

(b) Show that

$$\sum_{i=1}^{I}\sum_{a=1}^{A} b_a \frac{(\overline{e}_{ia} - \overline{e}_a - \overline{e}_i + \overline{e})^2}{\sigma^2} \sim \chi^2_{(A-1)(I-1)}.$$

(c) Similarly, obtain the distribution of $\sum_{i=1}^{I}\sum_{a=1}^{A} b_a(\varepsilon_{ia} - \overline{\varepsilon}_a)/\sigma^2_{\iota a}$ and hence deduce the expected value of $\sum_{i=1}^{I}\sum_{a=1}^{A} b_a(\overline{Y}_{ia} - \overline{Y}_i - \overline{Y}_a + \overline{Y})^2$.

(d) Similarly, obtain the expected values of the other sums of squares in constructing the ANOVA for the model.

7.9 Consider the model and ANOVA in Exercise 7.8.

(a) Derive the generalized p-value for testing null hypotheses of the form

$$H_0 : \sigma_i^2 \leq \sigma_0^2.$$

(b) Similarly establish a testing procedure for testing the null hypotheses of the form

$$H_0 : \sigma_{ia}^2 \geq \sigma_0^2.$$

(c) Deduce the left-sided $100\gamma\%$ confidence interval for σ_i^2.
(d) Deduce the equal-tail $100\gamma\%$ confidence interval for σ_{ia}^2.

7.10 Consider the following data reported by Crowder and Hand (1990) concerning a dietary treatment. Observations on one of the measurements (plasma ascorbic acid) were taken from 12 patients at 7 occasions, twice before, three times during, and twice after the treatment regime.

Patient reactions to treatment

	Phase I		Phase II			Phase III	
Patient\Week:	1	2	6	10	14	15	16
1	0.22	0.00	1.03	0.67	0.75	0.65	0.59
2	0.18	0.00	0.96	0.96	0.98	1.03	0.70
3	0.73	0.37	1.18	0.76	1.07	0.80	1.10
4	0.30	0.25	0.74	1.10	1.48	0.39	0.36
5	0.54	0.42	1.33	1.32	1.30	0.74	0.56
6	0.16	0.30	1.27	1.06	1.39	0.63	0.40
7	0.30	1.09	1.17	0.90	1.17	0.75	0.88
8	0.70	1.30	1.80	1.80	1.60	1.23	0.41
9	0.31	0.54	1.24	0.56	0.77	0.28	0.40
10	1.40	1.40	1.64	1.28	1.12	0.66	0.77
11	0.60	0.80	1.02	1.28	1.16	1.01	0.67
12	0.73	0.50	1.08	1.26	1.17	0.91	0.87

Assuming model (7.70) set up the ANOVA table. Test the hypothesis that there is no significant difference between treatment phases and discuss your findings. Also test the interaction between occasions and treatments. Test for the significance of variance components and construct 95% confidence intervals.

7.11 Consider the data set shown in Table 7.4 on the effects of three diet supplements. Carry out an RM ANCOVA to compare the diet supplements in a regression setting in which the vector of response means prior to administering the diets is a covariate in the regression model. Also test the effect of time and the interaction effects of diet supplements over time.

CHAPTER 8

REPEATED MEASURES UNDER HETEROSCEDASTICITY

8.1 INTRODUCTION

In the treatment of repeated measures in Chapter 7 we assumed that all treatment groups have equal variances. Traditionally, this assumption is made for simplicity and mathematical tractability. While there is no serious problem when the assumption is reasonable, as demonstrated by Ho, Weerahandi, and Hung (2002), the assumption can lead to serious erroneous conclusions when the variances are substantially different. Recall that the conventional ANOVA problems that rely on the equal variances assumption can dramatically reduce the power of tests. Moreover, the magnitude of the lack of power problem of classical ANOVA tests based on that assumption increases with the number of treatments being compared. Moreover, in situations of higher-way ANOVA under heteroscedasticity, one can even make misleading conclusions. For example, as we saw in Chapter 2, mislead by a classical F-test, one may conclude that a certain factor of an ANOVA is significant when in fact a different factor is significant. As we will see later in this chapter, tests in repeated measures context also suffer from similar drawbacks. Since even simple repeated measures models involve a number of factors, it is important that we do not

(Generalized Inference in Repeated Measures, Edition 2). By (Weerahandi)

make the assumption of equal variances for the sake of simplicity unless the assumption is very reasonable.

In view of these considerations, the purpose of this chapter is to establish procedures for making inferences about unknown parameters without the assumption of equal group variances. The treatment in this chapter will be limited to the most widely used two-factor model discussed in Chapter 7. Although it is not difficult to extend procedures to other types of design, this is an area requiring further research.

If the covariance matrices are unstructured, the problem is one in MANOVA under unequal covariances, a problem that we have already addressed in Chapter 6. If the covariance matrix is structured as specified in Chapter 7 and if the group variances are unequal, then there are no exact classical tests available for testing any of the hypotheses concerning fixed effects discussed in Chapter 7. Nevertheless, the problem can be tackled by taking the generalized approach as in other situations of heteroscedasticity. This is accomplished by considering the appropriate sums of squares in the case of known variances and handling the unknown variances by their estimates in such a way that the resulting probability statements become exact.

8.2 TWO-FACTOR MODEL WITH UNEQUAL GROUP VARIANCES

Consider the two-factor repeated measures model in which possibly unequal numbers of subjects are nested under the treatment groups. Again it should be emphasized that, in designed experiments involving repeated measures, among-subjects variation should be minimized by taking some observations from each subject before the treatments are given. Then, the corrected responses Y_i's are computed by subtracting the pre-treatment mean responses from the corresponding data observed after the treatments. We assume that response data from each subject has already been corrected in this manner whenever such pre-experimental data are available.

Suppose there are G groups and there are n_g subjects in group g. Let $\sum n_g = N$ be the total number of subjects used in the experiment. Each subject is observed at T equally or unequally spaced time points . Let $Y_{i(g)t}$ denote the value of the response variable (already corrected if pre-experimental data are available) taken at the tth time point from the ith subject in group g. Consider again the mixed model

$$Y_{i(g)t} = \theta_g + \beta_t + \gamma_{gt} + \alpha_{i(g)} + \epsilon_{i(g)t}; \tag{8.1}$$
$$t = 1, \ldots, T; i(g) = 1, \ldots, n_g; g = 1, \ldots, G$$

discussed in Chapter 7, but without the assumption of equal error variances. Recall that $\alpha_{i(g)}$ is the random effect due to among-subject variation, θ_g, $g = 1, \ldots, G$ are the treatment (or factor 1) effects, β_t, $t = 1, \ldots, T$ are effects due to occasions (or factor 2), γ_{gt} are their interactions, and ϵ_{it} are the residual terms. Extending the usual assumption about variance components with possibly unequal group variances, we now have

$$\alpha_{i(g)} \sim N(0, \sigma_\alpha^2), \quad \epsilon_{i(g)t} \sim N(0, \sigma_g^2); \tag{8.2}$$
$$t = 1, \ldots, T;\ i(g) = 1, \ldots,\ n_g; g = 1, \ldots, G.$$

As before, all time-dependent parameters can be represented in a single equation by rewriting model (8.1) in terms of the vector of all observations from subject $i(g)$, $\mathbf{Y}_{i(g)} = \left(Y_{i(g)1}, .., Y_{i(g)T}\right)'$. Then the model can be expressed as

$$\mathbf{Y}_{i(g)} = \mu \mathbf{1}_T + \delta_g \mathbf{1}_T + \boldsymbol{\beta} + \boldsymbol{\gamma}_g + \boldsymbol{\varepsilon}_{i(g)}, \tag{8.3}$$

where $\theta_g = \mu + \delta_g$, μ is the grand mean so that $\sum \delta_g = 0$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_g$ are $T \times 1$ vectors formed by β_t's and γ_{gt}'s, and

$$\boldsymbol{\varepsilon}_{i(g)} \sim N(\mathbf{0}, \Sigma_g), \quad \text{with } \Sigma_g = \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T' + \sigma_g^2 \mathbf{I}_T$$

8.3 POINT ESTIMATION

First consider the problem of estimating unknown parameters of the problem when the variance components have been specified or estimated. This will also prove to be important in developing the ANOVA based on an orthogonal decomposition of the total sums of squares. Point estimation of location parameters requires some constraints to make them estimable. To develop tractable distributions consider the natural constraints

$$\sum_{g=1}^{G} \left(n_g \Sigma_g^{-1} \boldsymbol{\delta}_g\right) = 0, \quad \mathbf{1}_T' \boldsymbol{\beta} = 0, \quad \mathbf{1}_T' \boldsymbol{\gamma}_g = \mathbf{0}\ \forall g, \tag{8.4}$$

and

$$\sum_{g=1}^{G} \left(n_g \Sigma_g^{-1} \boldsymbol{\gamma}_g\right) = \mathbf{0}.$$

Let

$$\begin{aligned} \overline{Y}_{gt} &= \frac{1}{n_g} \sum_{i \in g} Y_{i(g)t}; \\ g &= 1, \ldots, G\ ,\ t = 1, \ldots, T \end{aligned}$$

be the sample of mean n_g observations from subjects in group and let $\overline{Y}$ be the grand mean of all the data. Also define the $T \times 1$ vector

$$\overline{\mathbf{Y}}_g = \begin{pmatrix} \overline{Y}_{g1} \\ \overline{Y}_{g2} \\ \vdots \\ \overline{Y}_{gT} \end{pmatrix},$$

and various sample means, namely the group means, occasion means, and subject means, as

$$\overline{Y}_g = \frac{1}{n_g T} \sum_{i \in g} \sum_{t=1}^{T} Y_{i(g)t} = \frac{\mathbf{1}_T^{'} \overline{\mathbf{Y}}_g}{T} \ ,$$

$$\overline{Y}_t = \frac{1}{N} \sum_{g=1}^{G} \sum_{i \in g} Y_{i(g)t} \ ,$$

$$\overline{Y}_{i(g)} = \frac{1}{T} \sum_{t=1}^{T} Y_{i(g)t} \ .$$

It can be shown (Exercise 8.1) that the maximum likelihood estimates of the parameters under the constraints (8.4) are

$$\widehat{\mu} = \left(\sum_{g=1}^{G} \left(n_g \Sigma_g^{-1} \right) \right)^{-1} \sum_{g=1}^{G} \left(n_g \Sigma_g^{-1} \overline{\mathbf{Y}}_g \right) \mathbf{1}, \tag{8.5}$$

$$\hat{\boldsymbol{\beta}} = \left(\sum_{g=1}^{G} n_g \Sigma_g^{-1} \right)^{-1} \sum_{g=1}^{G} \left(n_g \Sigma_g^{-1} \left(\overline{\mathbf{Y}}_g - \overline{Y}_g \mathbf{1} \right) \right), \tag{8.6}$$

$$\widehat{\delta}_g = \overline{Y}_g - \widehat{\mu},$$

and

$$\widehat{\boldsymbol{\gamma}}_g = \overline{\mathbf{Y}}_g - \overline{Y}_g \mathbf{1} - \hat{\boldsymbol{\beta}} \tag{8.7}$$

for $g = 1, \ldots, G$. The covariance matrices are of course typically unknown and so the parameters are estimated after replacing the variance components by their estimates, which will be developed later in this chapter.

8.4 TESTING FIXED EFFECTS

Testing various hypotheses of interest by the generalized approach requires development of appropriate test variables. When the group error variances are unequal, we also need to work out the distribution of appropriately weighted sums of squares that comprise the test variables. Since it is the group variances that are different, the weighting needs to be applied only to the group sample means. Thus, the counterpart of (7.31) takes the form

$$\overline{Y}_{i(g)} - \overline{Y}_{\omega} = (\overline{Y}_g - \overline{Y}_{\omega}) + (\overline{Y}_{i(g)} - \overline{Y}_g), \tag{8.8}$$

where

$$\overline{Y}_g = \frac{1}{n_g T} \sum_{i \in g} \sum_{t=1}^{T} Y_{i(g)t},$$

and

$$\overline{Y}_\omega = \frac{\sum_{g=1}^{G} \omega_g \overline{Y}_g}{\sum_{g=1}^{G} \omega_g},$$

where

$$\omega_g = \frac{n_g}{\sigma_\alpha^2 + \sigma_g^2/T} \ .$$

Since

$$\sum_{i \in g} (\overline{Y}_{i(g)} - \overline{Y}_g) = 0$$

we get the weighted sums of squares decomposition

$$\sum_{g=1}^{G} \sum_{i \in g} w_g (\overline{Y}_{i(g)} - \overline{Y}_\omega)^2 = T \sum_{g=1}^{G} \omega_g (\overline{Y}_g - \overline{Y}_\omega)^2 + \sum_{g=1}^{G} \sum_{i \in g} w_g (\overline{Y}_{i(g)} - \overline{Y}_g)^2,$$

where $w_g = \omega_g / n_g$. The standardized between-group sum of squares

$$\widetilde{S}_b = T \sum_{g=1}^{G} \omega_g \left(\overline{Y}_{g.} - \overline{Y}_\omega \right)^2, \tag{8.9}$$

is distributed independently of

$$S_g = T \sum_{i \in g} (\overline{Y}_{i(g)} - \overline{Y}_g)^2, g = 1, \ldots, G,$$

which are also mutually independent as they are computed from subjects from different groups. To derive the distributions of these quantities, let us start with

$$\begin{aligned} \overline{Y}_{i(g)} &= \mu + \delta_g + \alpha_{i(g)} + \epsilon_{it} \\ &\sim N(\mu + \delta_g, \sigma_\alpha^2 + \sigma_g^2/T), \ i = 1, \ldots, \ n_g \ . \end{aligned}$$

Now it follows from distributional results involving a single sample from a normal population that

$$V_g = \frac{S_g}{T\sigma_\alpha^2 + \sigma_g^2} \sim \chi^2_{n_g - 1}, \ g = 1, \ldots, G \tag{8.10}$$

and in turn that

$$\widetilde{S}_{wg} = \sum_{g=1}^{G} \frac{S_g}{T\sigma_\alpha^2 + \sigma_g^2} \sim \chi^2_{N-G} \ . \tag{8.11}$$

Moreover, from known results on weighted least-squares regression, we get

$$W = \widetilde{S}_b \sim \chi^2_{G-1} \tag{8.12}$$

if the treatment means are equal.

8.4.1 Comparing Treatment Groups

Now we are in a position to develop procedure for testing various hypotheses of interest. First consider the null hypothesis that there is no difference between the treatment groups−that is, the hypothesis

$$H_{01} : \delta_1 = \delta_2 = \cdots = \delta_T = 0. \tag{8.13}$$

Tests of the hypothesis can be based on the distributions given in the previous section. In order to find an appropriate extreme region to base our test, write the standardized between group sum of squares as $\widetilde{S}_b = \widetilde{S}_b\left(\eta_1^2, ..., \eta_G^2\right)$, where $\eta_g^2 = \sigma_\alpha^2 + \sigma_g^2/T$. Let $\widetilde{s}_b()$ be the observed value of $\widetilde{S}_b(\ )$. Consider the potential extreme region,

$$C = \left\{\mathbf{Y} \mid \widetilde{S}_b\left(\eta_1^2, ..., \eta_G^2\right) \geq \widetilde{s}_b\left(\frac{\eta_1^2 s_1}{S_1}, ..., \frac{\eta_G^2 s_G}{S_G}\right)\right\}.$$

Using (8.10) and (8.12), its probability can be expressed as

$$p = \Pr(C) = \Pr\left\{W \geq \widetilde{s}_b\left(\frac{s_1}{V_1}, ..., \frac{s_G}{V_G}\right)\right\}, \tag{8.14}$$

where $V_1, \ldots, V_G$ have chi-squared distributions irrespective of the validity of the null hypothesis. Moreover, W has a central chi-squared distribution under the null hypothesis and a noncentral chi-squared distribution otherwise, leading to higher probability of C for any deviation from the null hypothesis. Hence, C is indeed an extreme region with its probability free of unknown parameters.

The probability can be easily computed by Monte Carlo method by generating a large number of random numbers from the chi-squared random variables and then finding the fraction of times the inequality in formula (8.14) is satisfied. The p-value can also be evaluated numerically for any desired level of accuracy. The accuracy of the integration could be enhanced by expressing it in terms of some beta random variables as we did in the one-way ANOVA problem. To see this and that the resulting test is a generalized F-test, express C as

$$\begin{aligned}
C &= \{\mathbf{Y} \mid \widetilde{S}_b \geq \widetilde{s}_b\left(\frac{s_1}{V_1}, ..., \frac{s_G}{V_G}\right)\} \\
&= \{\mathbf{Y} \mid \widetilde{S}_b \geq \widetilde{s}_b(\frac{s_1}{\widetilde{S}_{wg} B_1 B_2 ... B_{G-1}}, \\
&\qquad \frac{s_2}{\widetilde{S}_{wg}\left(1 - B_1\right) B_2 ... B_{G-1}}, ..., \frac{s_G}{\widetilde{S}_{wg}\left(1 - B_{G-1}\right)})\} \\
&= \{\mathbf{Y} \mid \frac{\widetilde{S}_b/(G-1)}{\widetilde{S}_{wg}/\left(N - G\right)} \\
&\geq \widetilde{s}_b(\frac{s_1}{B_1 B_2 ... B_{G-1}}, ..., \frac{s_G}{(1 - B_{G-1})})\frac{N-G}{G-1}\},
\end{aligned} \tag{8.15}$$

where

$$B_j = \frac{\left(\sum_{g=1}^{j} V_g\right)}{\left(\sum_{g=1}^{j+1} V_g\right)} \sim \text{Beta}\left(\sum_{g=1}^{j} \frac{n_g - 1}{2}, \frac{n_{j+1} - 1}{2}\right), j = 1, 2, ..., G - 1.$$

Under the null hypothesis

$$\frac{\widetilde{S}_b/(G-1)}{\widetilde{S}_{wg}/\,(N-G)} \sim F_{G-1,N-G} \tag{8.16}$$

and if the null hypothesis is not true, its distribution is noncentral F, leading to higher probability of 8.15 for any deviation from the null hypothesis. Therefore, the p-value of the generalized F-test given by (8.15) and (8.16) can be computed as

$$p = 1 - E\left\{H_{G-1,N-G}\left[\frac{N-G}{G-1}\widetilde{s}_b(\mathbf{s}, \mathbf{B})\right]\right\}, \tag{8.17}$$

where

$$\begin{aligned}\widetilde{s}_b(\mathbf{s}, \mathbf{B}) \quad = \quad & \widetilde{s}_b(\frac{s_1}{B_1 B_2 ... B_{G-1}}, \frac{s_2}{(1-B_1)\, B_2 ... B_{G-1}}, \\ & \frac{s_3}{(1-B_2)\, B_3 ... B_{G-1}}, ..., \frac{s_G}{(1-B_{G-1})}),\end{aligned}$$

$H_{G-1,N-G}$ is the cdf of the F distribution with $G-1$ and $N-G$ degrees of freedom and the expectation is taken with respect to the independent beta random variables B_j, $j = 1, 2, ..., G-1$. This p-value can be conveniently computed using the XPro software package.

Example 8.1. Comparing consumer demands in four regions

Consider the consumer demand for a certain food item, say eggs. In a demand comparison of four regions based on sales data from a sample of supermarkets, suppose a measure of weekly demand level is measured as

$$\text{Demand} = 10{,}000 \frac{\text{Revenue from sale of eggs}}{\text{Revenue from all food sales}}.$$

Shown in Table 8.1 is a hypothetical data set that was generated by simulating model (8.1).

Although the data in Table 8.1 seem typical in a repeated measures design, a closer look at the data reveals that the group variances in this case are substantially different. This is evident from Table 8.2, which displays the group means and the standard deviations (MLEs) of the data

Table 8.1 Weekly demand for eggs in four regions

Region	**Store**	**Week 1**	**Week 2**	**Week3**	**Week 4**	**Week 5**
1	1	21.537	39.204	37.492	32.372	33.039
1	2	22.17	33.595	24.768	25.751	32.081
1	3	24.493	16.593	34.038	29.993	24.567
1	4	30.573	40.248	30.95	34.713	26.04
1	5	33.854	25.044	40.569	41.867	27.177
2	6	27.716	32.056	30.528	28.785	26.312
2	7	28.037	28.818	29.461	29.951	31.377
2	8	29.798	29.971	33.017	32.217	30.152
2	9	28.571	29.333	29.227	28.022	31.023
2	10	31.56	28.745	29.111	30.391	30.798
3	11	33.614	31.844	31.065	33.607	30.432
3	12	30.431	33.592	33.983	32.162	33.9
3	13	32.567	31.845	29.857	33.534	34.418
3	14	32.809	32.651	31.198	33.122	34.17
3	15	32.738	34.003	30.785	33.19	34.96
4	16	29.672	34.541	36.911	30.857	23.868
4	17	30.534	25.244	29.061	33.071	34.601
4	18	24.456	35.638	31.371	29.08	38.951
4	19	30.821	27.201	26.615	36.217	27.297
4	20	29.889	32.117	37.363	27.285	33.96

Obviously, in this application it is not reasonable to assume that the variances are equal. But does it make any difference in our conclusions whether or not the assumption is reasonable? To examine this, let us first ignore the fact that variances are different and apply the classical ANOVA as usually done by most practitioners for the sake of simplicity. The ANOVA table obtained by applying formulas given in Chapter 7 for the case of homoscedastic variances is shown below.

ANOVA Table

Source	DF	SS	F-value	p-value
Regions	3	110.99	2.137	0.136
Within region	16	277.00		
Weeks	4	86.83	1.29	0.284
Regions × Weeks	12	135.05	0.668	0.775
Error	64	1078.08		
Total	99	0.94		

Table 8.2 Weekly demand summary statistics

Region	Week 1	Week 2	Week 3	Week 4	Week 5
Sample Means					
1	26.53	30.94	33.56	32.94	28.58
2	29.14	29.78	30.27	29.87	29.93
3	32.43	32.79	31.38	33.12	33.58
4	29.07	30.95	32.26	31.30	31.74
Standard Deviations					
1	4.86	8.97	5.46	5.36	3.37
2	1.40	1.22	1.46	1.44	1.85
3	1.06	0.89	1.38	0.52	1.61
4	2.35	4.07	4.26	3.12	5.42

According to the p-values appearing in the ANOVA table, none of the effects including the regional effect is significant. Now let us drop the equal variances assumption and retest the hypothesis that there is no difference in the mean demand for eggs in different regions. This can be accomplished by applying formula (8.14) or (8.17). The p-value for testing the difference in regions then become 0.0008. This means that the difference in regional demand is highly significant despite what the classical ANOVA suggested. Usually milder assumptions make the p-value of a test larger and power of a test smaller. But here the assumption of equal variances is so unreasonable that the p-value under the assumption of equal variances is substantially larger. This example clearly demonstrates the tremendous deterioration of the power of classical F-tests under heteroscedasticity. Ho, Weerahandi, and Hung (2002) provide additional examples of this nature.

8.4.2 Testing the Interactions

Next consider the problem of testing the group $\times$ time interaction,

$$H_0 : \gamma_{gt} = 0 \quad \forall\, g = 1, .., G, t = 1, .., T.$$

As discussed in Ho, Weerahandi and Hung (2002), this hypothesis can be based on the normalized version of the corresponding sum of squares used in the classical ANOVA table,

$$\widetilde{S}_{og} = \widetilde{S}_{og}\left(\widehat{\theta}_g, \widehat{\boldsymbol{\beta}}; \Sigma_g, g=1,...,G\right) = \sum_{g=1}^{G} n_g^{'} \widehat{\boldsymbol{\gamma}}_g \Sigma_g^{-1} \widehat{\boldsymbol{\gamma}}_g \tag{8.18}$$

$$= \sum_{g=1}^{G} n_g \left(\overline{\mathbf{Y}}_g - \widehat{\theta}_g \mathbf{1}_T - \widehat{\boldsymbol{\beta}}\right)' \Sigma_g^{-1} \left(\overline{\mathbf{Y}}_g - \widehat{\theta}_g \mathbf{1}_T - \widehat{\boldsymbol{\beta}}\right) \tag{8.19}$$

Under H_0, from the classical ANOVA we get

$$\widetilde{S}_{og} \sim \chi^2_{(G-1)(T-1)}.$$

To handle the unknown variances case, rewrite $\widetilde{S}_{og}$ as $\widetilde{S}_{og}\left(\sigma_1^2, ..., \sigma_G^2; \delta_1^2, ..., \delta_G^2\right)$, where $\delta_g^2 = T\sigma_\alpha^2 + \sigma_g^2$. Notice that

$$\begin{aligned} \Sigma_g^{-1} &= \frac{1}{\sigma_g^2} I - \frac{\sigma_\alpha^2}{\sigma_g^2 \delta_g^2} \mathbf{1}\mathbf{1}' \\ &= \frac{1}{\sigma_g^2} I - \frac{\left(\delta_g^2 - \sigma_g^2\right)/T}{\sigma_g^2 \delta_g^2} \mathbf{1}\mathbf{1}' \end{aligned}$$

is a function of only σ_g^2 and δ_g^2. From known results on the equal variances case, which are still valid within each group, we get

$$U_g = \frac{S_{eg}}{\sigma_g^2} \sim \chi^2_{(T-1)(n_g-1)} \tag{8.20}$$

and

$$V_g = \frac{S_g}{\delta_g^2} \sim \chi^2_{n_g-1}, \tag{8.21}$$

where

$$S_{eg} = \sum_{t=1}^{T} \sum_{i \in g} \left(Y_{i(g)t} - \overline{Y}_{i(g)} - \overline{Y}_{gt} + \overline{Y_g}\right)^2$$

and

$$S_g = T \sum_{i \in g} \left(\overline{Y}_{i(g)} - \overline{Y}_g\right)^2.$$

As in the previous section, we can find a generalized test using the extreme region

$$\begin{aligned} C = \{\mathbf{Y} \mid \widetilde{S}_{og}\left(\sigma_1^2, ..., \sigma_G^2; \delta_1^2, ..., \delta_G^2\right) \geq \widetilde{s}_{og}(\frac{\sigma_1^2 s_{e1}}{S_{e1}}, ..., \frac{\sigma_G^2 s_{eG}}{S_{eG}}; \\ \frac{\delta_1^2 s_1}{S_1}, ..., \frac{\delta_G^2 s_G}{S_G})\}, \end{aligned} \tag{8.22}$$

where lowercase letters stand for observed values of the random variables. Clearly the observed sample points fall on the boundary of this subset of the sample space and its probability increases for any departure from H_{02}. Hence, C is an extreme region leading to generalized tests. Its p-value can be computed as

$$\begin{aligned} p &= 1 - \Pr[\widetilde{S}_{og}\left(\sigma_1^2, ..., \sigma_G^2; \delta_1^2, ..., \delta_G^2\right) \\ &\leq \widetilde{s}_{og}\left(\frac{\sigma_1^2 s_{e1}}{S_{e1}}, ..., \frac{\sigma_G^2 s_{eG}}{S_{eG}}; \frac{\delta_1^2 s_1}{S_1}, ..., \frac{\delta_G^2 s_G}{S_G}\right)] \\ &= 1 - E\left\{F_{\chi^2_{(G-1)(T-1)}}\left(\widetilde{s}_{og}\left(\frac{s_{e1}}{U_1}, ..., \frac{s_{eG}}{U_G}; \frac{s_1}{V_1}, ..., \frac{s_G}{V_G}\right)\right)\right\}, \quad (8.23) \end{aligned}$$

where the expectation is taken with respect to the independent chi-squared random variables U_g, V_g, $g = 1, .., G$ defined by equation (8.20).

8.4.3 Testing the equality of occasion effects

Finally consider the problem of testing the equality of occasion effects; i.e., H_0: $\boldsymbol{\beta} = \mathbf{0}$. Following the above approach, tests of H_0 can be based on the normalized occasion sum of squares $\widetilde{S}_o\left(\sigma_1^2, ..., \sigma_G^2; \delta_1^2, ..., \delta_G^2\right) = \widetilde{S}_o$, where

$$\begin{aligned} \widetilde{S}_o &= \sum_{g=1}^{G} n_g \widehat{\boldsymbol{\beta}}' \Sigma_g^{-1} \widehat{\boldsymbol{\beta}} \\ &= \sum_{g=1}^{G} n_g \left(\overline{\mathbf{Y}}_g - \widehat{\theta}_g \mathbf{1}_T - \widehat{\boldsymbol{\gamma}}_g\right)' \Sigma_g^{-1} \left(\overline{\mathbf{Y}}_g - \widehat{\theta}_g \mathbf{1}_T - \widehat{\boldsymbol{\gamma}}_g\right). \quad (8.24) \end{aligned}$$

Under H_0, we have

$$\widehat{S}_o^2 \backsim \chi^2_{T-1}. \quad (8.25)$$

By taking the approach in previous sections, the extreme region is now defined as

$$\begin{aligned} C &= \{\mathbf{Y} \mid \widetilde{S}_o\left(\sigma_1^2, ..., \sigma_G^2; \delta_1^2, ..., \delta_G^2\right) \geq \widetilde{s}_o(\frac{\sigma_1^2 s_{e1}}{S_{e1}}, ..., \frac{\sigma_G^2 s_{eG}}{S_{eG}}; \\ &\quad \frac{\delta_1^2 s_1}{S_1}, ..., \frac{\delta_G^2 s_G}{S_G})\}. \quad (8.26) \end{aligned}$$

Its p-value is computed as

$$\begin{aligned} p &= 1 - \Pr[\widetilde{S}_o\left(\sigma_1^2, ..., \sigma_G^2; \delta_1^2, ..., \delta_G^2\right) \\ &\leq \widetilde{s}_o\left(\frac{\sigma_1^2 s_{e1}}{S_{e1}}, ..., \frac{\sigma_G^2 s_{eG}}{S_{eG}}; \frac{\delta_1^2 s_1}{S_1}, ..., \frac{\delta_G^2 s_G}{S_G}\right)] \\ &= 1 - EF_{\chi^2_{T-1}}\left(\widehat{s}_o^2\left(\frac{s_{e1}}{U_1}, ..., \frac{s_{eG}}{U_G}; \frac{s_1}{V_1}, ..., \frac{s_G}{V_G}\right)\right), \quad (8.27) \end{aligned}$$

where the expectation is taken with respect to the independent chi-squared random variables $U_g, V_g, g = 1, .., G$ defined by equation (8.20). The p-values (8.23) and (8.27) can be computed by exact (up to desired level of accuracy) numerical integration or can be well estimated by Monte Carlo integration. In the later case, both probabilities can be computed simultaneously using the same set of chi-squared random variates generated from $U_g, V_g, g = 1, .., G$.

Example 8.2. Comparing consumer demands in four regions (continued).

Consider again the data set in Table 8.1. Now we are in a position to test the null hypothesis of constant demand for eggs over the study period and the interaction between the time and region without relying on the assumption of equal variances. Application of formula (8.27) with current data yields a p-value of 0.37 for testing the equality of week effects. So, there is no reason to doubt the null hypothesis of constant weekly demand. From (8.23) we get a p-value of 0.64, suggesting that there is no interaction between the weeks and the regions. These are the same conclusions we made based on the classical ANOVA, but now we are not relying on the assumption of equal variances.

8.5 MULTIPLE COMPARISONS

Now suppose we have established the significance of difference in factor effects so that we can proceed to do multiple comparisons. Comparing treatment groups is of special importance. As pointed out in Chapter 7, except in some special cases, specialized approaches such as the Tukey-type methods and Scheffe-type methods fail in repeated measures problems even in the case of homoscedastic group variances. Nevertheless, the Bonferroni method, which is based on pair-wise comparisons, still allows us to carry out multiple comparisons. Since the Bonferroni procedures are usually more powerful than the Scheffe procedures in most practical applications involving reasonable number of multiple comparisons, we can take the Bonferroni approach to tackle the current problem without sacrificing the power of the test.

First consider the problem of comparing two treatment groups, say g_1 and g_2. Recall that the point estimate of the mean effect due to group g without any grand mean correction is $\widehat{\theta}_g = \overline{Y}_g$. We can easily obtain its distribution as

$$\begin{aligned} \overline{Y}_g &= \theta_g + \overline{\alpha}_g + \overline{\epsilon}_g \\ &\sim N(\theta_g, \frac{1}{Tn_g}(T\sigma_\alpha^2 + \sigma_g^2), \end{aligned}$$

Let $g1$ and $g2$ be the two groups of interest. Then the difference in sample group means is distributed as

$$\overline{Y}_{g1} - \overline{Y}_{g2} \sim N(\theta_{g1} - \theta_{g2}, \frac{1}{T}(\frac{\delta_{g_1}^2}{n_{g1}} + \frac{\delta_{g_2}^2}{n_{g2}}), \tag{8.28}$$

where $\delta_g^2 = T\sigma_\alpha^2 + \sigma_g^2$. Since $\delta_{g_1}^2$ and $\delta_{g_1}^2$ are nuisance parameters, we can tackle them as in the Behrens−Fisher problem using the within group sample variance S_{g_1} and S_{g_2}, which are unbiased estimates of the nuisance parameters. Recall that the distribution of within group sample variance is given by

$$V_g \equiv \frac{S_g}{\delta_g^2} \sim \chi^2_{n_g-1}. \tag{8.29}$$

8.5.1 Hypothesis testing

To illustrate the approach we can take in hypothesis testing, consider one-sided hypotheses of the form

$$H_0 : \theta_{g_1} - \theta_{g_2} \leq \theta_0 \tag{8.30}$$

for comparing the two groups of interest. By taking the approach of Tsui and Weerahandi (1989) to derive the solution to the Behrens−Fisher problem in a formal manner, or by the substitution method discussed in Chapter 1, we get the appropriate extreme region for testing H_0 as

$$\begin{aligned} C &= \left\{ \mathbf{Y} \mid \frac{(\overline{Y}_{g1} - \overline{Y}_{g2}) - \theta_0}{\sqrt{\frac{1}{T}(\frac{\delta_{g1}^2}{n_{g1}} + \frac{\delta_{g2}^2}{n_{g2}})}} \geq \frac{\overline{y}_{g1} - \overline{y}_{g2} - \theta_0}{\sqrt{\frac{1}{T}(\frac{s_{g1}}{V_{g_1} n_{g1}} + \frac{s_{g2}}{V_{g_2} n_{g2}})}} \right\} \\ &= \left\{ \mathbf{Y} \mid Z \geq \frac{\overline{y}_{g1} - \overline{y}_{g2} - \theta_0}{\sqrt{\frac{1}{T}(\frac{s_{g1}}{V_{g_1} n_{g1}} + \frac{s_{g2}}{V_{g_2} n_{g2}})}} \right\}, \end{aligned}$$

where, when $\theta_{g_1} - \theta_{g_2} = \theta_0$, Z is a standard normal random variable. Now it is evident that the generalized p-value for testing left-sided hypotheses of the form (8.30) is given by (see Exercise 8.3)

$$\begin{aligned} p &= \Pr \left\{ Z \geq \frac{\overline{y}_{g1} - \overline{y}_{g2} - \theta_0}{\sqrt{\frac{1}{T}(\frac{s_{g1}}{V_{g_1} n_{g1}} + \frac{s_{g2}}{V_{g_2} n_{g2}})}} \right\} \\ &= 1 - EG_{n_{g_1}+n_{g_2}-2} \left[\frac{(\overline{y}_{g1} - \overline{y}_{g2} - \theta_0)\sqrt{n_{g_1} + n_{g_2} - 2}}{\sqrt{\frac{1}{T}(\frac{s_{g1}}{B n_{g1}} + \frac{s_{g2}}{(1-B) n_{g2}})}} \right], \end{aligned} \tag{8.31}$$

where $G_{n_{g_1}+n_{g_2}-2}$ is the cumulative distribution function of the Student's t distribution with $n_{g_1} + n_{g_2} - 2$ degrees of freedom and the expectation is with respect to the beta random variable

$$B = \frac{V_{g_1}}{V_{g_1} + V_{g_2}} \sim \text{Beta}(\frac{n_{g_1} - 1}{2}, \frac{n_{g_2} - 1}{2}).$$

This is the repeated measures counterpart of the generalized t-test given by Tsui and Weerahandi (1989). In fact the p-value can be computed by invoking any software that provide the generalized t-test to the Behrens−Fisher problem.

8.5.2 Confidence intervals

Next consider the problem of constructing confidence intervals for the difference in the effects of two treatment groups, say $g1$ and $g2$. Let $\theta = \theta_{g_1} - \theta_{g_2}$ be the parameter of interest. Generalized intervals could be derived using a generalized pivotal quantity or deduced simply from the generalized p-value given above. The latter amounts to finding one-sided $100\gamma\%$ generalized confidence intervals based on the probability statement

$$\Pr\left\{ Z \leq \frac{\overline{y}_{g1} - \overline{y}_{g2} - \theta}{\sqrt{\frac{1}{T}\left(\frac{s_{g_1}}{V_{g_1} n_{g1}} + \frac{s_{g2}}{V_{g_2} n_{g2}}\right)}} \right\} = \gamma. \tag{8.32}$$

It immediately follows from (8.32) that the $100\gamma\%$ generalized confidence for θ is

$$\theta \geq (\overline{y}_{g1} - \overline{y}_{g2}) - k_\gamma(s_{g_1}, s_{g2}),$$

where $k_\gamma = k_\gamma(s_{g_1}, s_{g2})$ is chosen such that

$$EG_{n_{g_1}+n_{g_2}-2}\left[\frac{k_\gamma \sqrt{n_{g_1} + n_{g_2} - 2}}{\sqrt{\frac{1}{T}\left(\frac{s_{g1}}{B n_{g1}} + \frac{s_{g2}}{(1-B) n_{g2}}\right)}} \right] = \gamma. \tag{8.33}$$

Other types of generalized confidence intervals can also be constructed in a similar manner. Of particular importance are the two-sided intervals. It can be deduced from one-sided intervals that the $100\gamma\%$ equal-tail generalized confidence for θ is

$$(\overline{y}_{g1} - \overline{y}_{g2}) - k(s_{g_1}, s_{g2}) \leq \theta \leq (\overline{y}_{g1} - \overline{y}_{g2}) + k(s_{g_1}, s_{g2}), \tag{8.34}$$

where $k = k(s_{g_1}, s_{g2}) = k_{(1+\gamma)/2}$ is computed by replacing γ in equation (8.33) by $(1+\gamma)/2$.

If there were r prespecified pairwise comparisons of interest, simultaneous confidence intervals can be obtained by applying foregoing formulas with the Bonferroni adjustment. For example, in constructing simultaneous generalized confidence intervals for r pairs of differences in group means, we apply (8.34) with α/r in place of $\alpha = 1 - \gamma$. This means that the $100\gamma\%$ simultaneous interval for a particular pair of means, say $\theta = \theta_{g_1} - \theta_{g_2}$, is computed as

$$(\overline{y}_{g1} - \overline{y}_{g2}) - k_r(s_{g_1}, s_{g2}) \leq \theta \leq (\overline{y}_{g1} - \overline{y}_{g2}) + k_r(s_{g_1}, s_{g2}),$$

where $k_r = k_{r,\gamma}(s_{g_1}, s_{g2})$ is chosen such that

$$EG_{n_{g_1}+n_{g_2}-2}\left[\frac{k_r\sqrt{n_{g_1}+n_{g_2}-2}}{\sqrt{\frac{1}{T}(\frac{q_{g_1}}{Bn_{g1}}+\frac{q_{g_2}}{(1-B)n_{g2}})}}\right] = 1-\frac{\alpha}{2r}$$
$$= 1-\frac{1-\gamma}{2r}, \qquad (8.35)$$

where q_g is the observed value of

$$Q_g = \frac{S_g}{Tn_g} = \frac{\sum\limits_{i\in g}(\overline{Y}_{i(g)}-\overline{Y}_g)^2}{n_g}.$$

Example 8.3. Comparing consumer demands in four regions (continued)

Consider again the data set in Table 8.1. Suppose we wish to compare the mean demand is region 1 with each of the other three groups. The table below shows the group means and the sums of squared deviation from store means for each of the four regions on which we can base our comparison.

Group	Group Mean	S_g	Q_g
1	30.51	240.28	12.014
2	29.80	12.65	0.633
3	32.66	3.06	0.152
4	31.06	21.01	1.051

With this summary information or directly using $\overline{Y}_{i(g)}$ raw data, generalized intervals for mean differences can be constructed using a software package such as XPro that provide solutions to the Behrens−Fisher problem. In constructing simultaneous confidence intervals, say at 95% level, the parameter α/r should be set as $0.05/3 = 0.0167$. This means that, in order the simultaneous confidence level to be 95%, we should construct pairwise intervals at 98.33%. The table below shows the 95% generalized simultaneous confidence intervals obtained in this manner. It is evident from the confidence intervals that the difference between θ_1 and θ_3 are statistically significant and that θ_1 is not significantly different from θ_2 and θ_4.

Mean Difference	Estimate	Conf. Interval
$\theta_1 - \theta_2$	0.71	-1.15 , 2.57
$\theta_3 - \theta_1$	2.15	0.32, 3.98
$\theta_4 - \theta_1$	0.55	-1.40, 2.38

8.6 INFERENCE ON VARIANCE COMPONENTS

Inferences on the group error variances σ_g^2, $g = 1, .., G$ are easily made using the chi-squared statistics

$$U_g = \frac{S_{eg}}{\sigma_g^2} \sim \chi^2_{(T-1)(n_g-1)}. \tag{8.36}$$

This implies, in particular, that $S_{eg}/(T-1)(n_g-1)$ is an unbiased estimate of σ_g^2. Inferences on the among-subject variance σ_α^2 is not straightforward although necessary distributional results are readily available from the result

$$V_g = \frac{S_g}{T\sigma_\alpha^2 + \sigma_g^2.} \sim \chi^2_{n_g-1}; \qquad g = 1, .., G. \tag{8.37}$$

Since the classical approach does not provide a systematic way of attacking the problem, here again we need to take the generalized approach. To do so, let

$$V \equiv \sum_{g=1}^{G} V_g \sim \chi^2_{N-G}.$$

Now consider, for instance, the null hypothesis,

$$H_0 : \sigma_\alpha^2 \leq \sigma_0^2, \tag{8.38}$$

Consider the extreme region given by

$$C = \left\{ \sum_{g=1}^{G} \frac{S_g}{T\sigma_\alpha^2 + \sigma_g^2} \geq \sum_{g=1}^{G} \frac{s_g}{T\sigma_\alpha^2 + \sigma_g^2 s_{eg}/S_{eg}} \right\} \tag{8.39}$$

$$= \left\{ V \geq \sum_{g=1}^{G} \frac{s_g}{T\sigma_\alpha^2 + \frac{s_{eg}}{U_g}} \right\}. \tag{8.40}$$

It is easily verified that C has all the required properties of an extreme region appropriate for defining generalized p-values. The resulting p-value is computed as

$$p = \underset{\sigma_\alpha^2}{Max} \Pr(C) = 1 - E\left[F_{\chi^2_{N-G}} \left(\sum_{g=1}^{G} \frac{q_g}{T\sigma_0^2 + s_{eg}/U_g} \right) \right], \tag{8.41}$$

where the expectation is taken with respect to the chi-squared random variables U_g, $g = 1, ..., G$, and $F_{\chi^2_{N-G}}$ is the cdf of the chi-squared distribution with $N - G$ degrees of freedom.

The generalized confidence intervals could be derived by using a generalized pivotal or deduced from the above p-value. In particular, the one-sided $100\gamma\%$ generalized confidence limits, say σ_0^2, for σ_α^2 are obtained from

$$E\left[F_{\chi^2_{N-G}}\left(\sum_{g=1}^{G}\frac{q_g}{T\sigma_0^2+s_{eg}/U_g}\right)\right]=k, \tag{8.42}$$

where $k=\gamma$ for the lower confidence limit and $k=1-\gamma$ for the upper confidence limit.

8.7 RM ANCOVA UNDER HETEROSCEDASTICITY

ANCOVA results presented in Chapter 7 could be easily extended to the case of unequal group variances. To outline the approach, as in that chapter, first consider the problem of testing the overall effect of the treatment during the trial period when the dependent variable of the regression is the average of responses taken over time. The results presented here also applies when one needs to estimate and compare the treatment effects for each time period separately, without MANOVA and without making any assumption on the correlation structure for the responses taken over time.

As before, to compare G treatment groups, consider the linear model

$$Y_{i(g)}=\theta_g+\sum_{k=1}^{K}\beta_k X_{ki(g)}+\epsilon_{i(g)}; \tag{8.43}$$

$$i(g)=1,\ldots,\ n_g; g=1,\ldots,G,$$

where $X_{ki(g)}$, $k=1,\ldots,G$ is a set of covariates. In Chapter 7 we assumed that the group variances are all equal. Now let us drop that assumption and just assume that their error terms independently and normally distributed as

$$\epsilon_{i(g)}\sim N(0,\sigma_g^2). \tag{8.44}$$

Results for this case can be deduced from that of Ananda (1998). To present the main results from Ananda (1998), let us continue to use the notation

$$\mathbf{y}'_g=\left(Y_{g1},Y_{g2},...,Y_{gn_g}\right)$$

to represent the responses available from Group g and let $\mathbf{y}=(\mathbf{y}'_1,\mathbf{y}'_2,...,\mathbf{y}'_G)'$ be the vector of all $N=\sum n_g$ responses. Let $\mathbf{X}_g$ denote the $n_{g.}\times K$ matrix of covariate data for the subjects in Group g and $\mathbf{X}=(\mathbf{X}'_1,\mathbf{X}'_2,...,\mathbf{X}'_G)'$ is a matrix of dimension $N\times K$. In matrix notation, the model can be expressed as

$$\begin{aligned}\mathbf{y} &= \mathbf{Z}\boldsymbol{\theta}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon} &(8.45)\\ &= \mathbf{W}\boldsymbol{\lambda}+\boldsymbol{\epsilon}, &(8.46)\end{aligned}$$

where

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}),$$

$$\boldsymbol{\theta} = (\theta_1, \theta, \ldots, \theta_G)', \ \boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_K)', \ \boldsymbol{\lambda} = (\boldsymbol{\theta}, \boldsymbol{\beta})',$$

$$\mathbf{Z} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1}_{n_G} \end{pmatrix}_{N \times G}$$

is the design matrix formed by dummy variables indicating the groups,

$$\mathbf{W} = (\mathbf{Z}, \mathbf{X}),$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 \mathbf{I}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_{n_2} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \sigma_G^2 \mathbf{I}_{n_G} \end{pmatrix}_{N \times N.}$$

is the variance covariance matrix. Being a block diagonal matrix, $\boldsymbol{\Sigma}$ can also be expressed as a direct sum of matrices as

$$\boldsymbol{\Sigma} = \sigma_1^2 \mathbf{I}_{n_1} \oplus \sigma_2^2 \mathbf{I}_{n_2} \oplus \cdots \cdot \oplus \sigma_G^2 \mathbf{I}_{n_G}.$$

Consider the problem of testing the null hypothesis

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_G. \tag{8.47}$$

If the null hypothesis is true, then we have

$$\begin{aligned} \mathbf{y} &= \theta + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &= \mathbf{V}\widetilde{\boldsymbol{\beta}}, \end{aligned} \tag{8.48}$$

where θ is the common parameter under the null hypothesis, $\mathbf{V} = (\mathbf{1}_N, \mathbf{X})$, and $\widetilde{\boldsymbol{\beta}} = (\theta, \boldsymbol{\beta}')'$. If group variances were known, $\boldsymbol{\Sigma}$ becomes a known matrix and so we could base a test on the chi-squared statistic

$$\mathbf{y}'(\boldsymbol{\Sigma}^{-1}\mathbf{W}(\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\Sigma}^{-\mathbf{1}} - \boldsymbol{\Sigma}^{-\mathbf{1}}\mathbf{V}(\mathbf{V}'\boldsymbol{\Sigma}^{-1}\mathbf{V})^{-1}\mathbf{V}\boldsymbol{\Sigma}^{-\mathbf{1}})\mathbf{y} \sim \boldsymbol{\chi}^2_{G-1}. \tag{8.49}$$

Since the group variances σ_g^2 are in fact unknown parameters, they can be estimated by the residual sums of squares S_g^2 obtained by regressing $\mathbf{y}_g$ on $\theta g \mathbf{1}_{n_g} + \mathbf{X}_g$. Moreover, it is known from the standard regression theory that

$$U_g = \frac{S_g^2}{\sigma_g^2} \sim \boldsymbol{\chi}^2_{n_g - K - 1}. \tag{8.50}$$

It is now clear that ANCOVA under unequal group variances could be performed by taking the generalized approach. The generalized test can be obtained by the substitution method. As follows from Ananda (1998), the generalized p-value for testing H_0 is

$$p = 1 - E\left\{F_{\chi}\left(\mathbf{y}'(\mathbf{DW}(\mathbf{W}'\mathbf{DW})^{-1}\mathbf{W}'\mathbf{D} - \mathbf{DV}(\mathbf{V}'\mathbf{DV})^{-1}\mathbf{VD})\mathbf{y}\right)\right\}, \quad (8.51)$$

where

$$\mathbf{D} = \frac{U_1}{s_1^2}\mathbf{I}_{n_1} \oplus \frac{U_2}{s_2^2}\mathbf{I}_{n_2} \oplus \cdots \oplus \frac{U_G}{s_G^2}\mathbf{I}_{n_G}$$

F_{χ} is the cdf of the chi-squared distribution with $G-1$ degrees of freedom, s_g^2 is the observed value of S_g^2, and the expectation is taken with respect to the chi-squared random variables U_g, $g = 1, \ldots, G$. The generalized p-value is easily computed by Monte Carlo method by generating a large number of sets of random numbers from the chi-squared random variables, computing $F_{\chi}\left(\mathbf{y}'(\mathbf{DW}(\mathbf{W}'\mathbf{DW})^{-1}\mathbf{W}'\mathbf{D} - \mathbf{DV}(\mathbf{V}'\mathbf{DV})^{-1}\mathbf{VD})\mathbf{y}\right)$ for each set, and then estimating the expected value appearing in (8.51) by their average. Ananda (1998) also expressed (8.51) as a generalized F-test, in which the expectation is taken with respect to a set of beta random variables and the F_{χ} in (8.51) is replaced by the cdf of the F distribution with $G-1$ and $N-G-K$ degrees of freedom, just like in the classical ANCOVA.

8.7.1 Case of repeated measures

Now suppose we have repeated measures taken over time from each subject as often the case. Also assume that the matrix $\mathbf{X}$ of covariates is measured only once and that regression coefficients are the same for all groups, as assumed in Chapter 7. Then in place of (8.43) consider the model

$$\begin{aligned} Y_{i(g)t} &= \theta_g + \lambda_t + \gamma_{gt} + \sum_{k=1}^{K}\beta_k X_{ki(g)} + \varepsilon_{i(g)t}; \qquad (8.52) \\ t &= 1, \ldots, T;\ i(g) = 1, \ldots,\ n_g; g = 1, \ldots, G. \end{aligned}$$

where θ_g is the mean effect of treatment g, λ_t is the occasion effect at time t, γ_{gt} is their interaction, $\varepsilon_{i(g)t} = \alpha_{i(g)} + \epsilon_{i(g)t}$, and $\varepsilon_{i(g)} = (\varepsilon_{i(g)1}, \ldots, \varepsilon_{i(g)T})'$ is distributed as $\varepsilon_{i(g)} \sim N(0, \Sigma_g)$, where for each group g the error covariance matrix Σ_g has the compound symmetric covariance structure

$$\Sigma = \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T' + \sigma_g^2 \mathbf{I}_T.$$

A major advantage of a covariance matrix having the compound symmetric structure is that the generalized least-squares estimates (also the MLEs) of parameters are the same as the ordinary least-squares estimates. Also assume that λ_t and γ_{gt} satisfy the constraints given in Section 8.2. Then, as in that section, the treatment groups can be compared and the coefficients of covariates can be estimated on the data

$$\overline{Y}_{i(g)} = \theta_g + \sum_{k=1}^{K} \beta_k X_{ki(g)} + \overline{\varepsilon}_{i(g)} \tag{8.53}$$

$$= \theta_g + \mathbf{X}'_{i(g)}\boldsymbol{\beta} + \overline{\varepsilon}_{i(g)}, \tag{8.54}$$

where $\mathbf{X}'_{i(g)} = (X_{1i(g)}, X_{2i(g)}, ..., X_{Ki(g)})$,

$$\overline{Y}_{i(g)} = \frac{1}{T}\sum_{t=1}^{T} Y_{i(g)t}, \text{ and } \overline{\varepsilon}_{i(g)} = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i(g)t} \sim N(0, \sigma_\alpha^2 + \frac{\sigma_g^2}{T}).$$

As before, the model can be expressed in matrix notation as

$$\overline{\mathbf{y}} = \mathbf{Z}\boldsymbol{\theta} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$= \mathbf{W}\boldsymbol{\lambda} + \boldsymbol{\epsilon}, \tag{8.55}$$

where

$$\boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}\right),$$

$\overline{\mathbf{y}} = (\overline{\mathbf{y}}'_1, \overline{\mathbf{y}}'_2, ..., \overline{\mathbf{y}}'_G)'$, $\overline{\mathbf{y}}'_g = \left(\overline{Y}_{g1}, \overline{Y}_{g2}, ..., \overline{Y}_{gn_g}\right)$, and

$$\widetilde{\boldsymbol{\Sigma}} = \widetilde{\sigma}_1^2 \mathbf{I}_{n_1} \oplus \widetilde{\sigma}_2^2 \mathbf{I}_{n_2} \oplus \cdots \oplus \widetilde{\sigma}_G^2 \mathbf{I}_{n_G},$$

where $\widetilde{\sigma}_g^2 = \sigma_\alpha^2 + \sigma_g^2/T$. Now it is clear that the generalized p-value for testing the null hypothesis H_0 can be deduced by replacing $\mathbf{y}$ in (8.51) by $\overline{\mathbf{y}}$ and replacing σ_g^2 by $\widetilde{\sigma}_g^2$. As before, $\widetilde{\sigma}_g^2$ can be tackled by the error sum of squares $\widetilde{S}_g^2$ obtained by regressing $\overline{\mathbf{y}}_g$ on $\theta g \mathbf{1}_{n_g} + \mathbf{X}_g$. Hence, the generalized p-vale for testing the equality of treatment means can be computed as

$$p = 1 - E\left\{F_\chi\left(\overline{\mathbf{y}}'(\mathbf{DW}(\mathbf{W}'\mathbf{DW})^{-1}\mathbf{W}'\mathbf{D} - \mathbf{DV}(\mathbf{V}'\mathbf{DV})^{-1}\mathbf{VD})\overline{\mathbf{y}}\right)\right\}, \tag{8.56}$$

where

$$\mathbf{D} = \frac{U_1}{\widetilde{s}_1^2}\mathbf{I}_{n_1} \oplus \frac{U_2}{\widetilde{s}_2^2}\mathbf{I}_{n_2} \oplus \cdots \oplus \frac{U_G}{\widetilde{s}_G^2}\mathbf{I}_{n_G},$$

F_χ is the cdf of the chi-squared distribution with $G-1$ degrees of freedom, $\widetilde{s}_g^2$ is the observed value of $\widetilde{S}_g^2$, and the expectation is taken with respect to chi-squared random variables U_g, $g = 1, \ldots, G$.

Inferences on occasion means and interactions could also be obtained using the results established in this chapter, and they are left as an exercise.

Exercises

8.1 Consider the repeated measures model (8.1) and assume the constraints defined by (8.4). If the covariance are known, show that

$$\begin{aligned}\widehat{\mu} &= \left(\sum_{g=1}^{G}\left(n_g\Sigma_g^{-1}\right)\right)^{-1}\sum_{g=1}^{G}\left(n_g\Sigma_g^{-1}\overline{\mathbf{Y}}_g\right),\ \widehat{\delta}_g = \overline{Y}_g - \widehat{\mu}\ ,\\ \widehat{\boldsymbol{\beta}} &= \left(\sum_{g=1}^{G}n_g\Sigma_g^{-1}\right)^{-1}\sum_{g=1}^{G}\left(n_g\Sigma_g^{-1}\left(\overline{\mathbf{Y}}_g - \overline{Y}_g\mathbf{1}\right)\right),\end{aligned}$$

and

$$\widehat{\boldsymbol{\gamma}}_g = \overline{\mathbf{Y}}_g - \overline{Y}_{g.}\mathbf{1} - \widehat{\boldsymbol{\beta}}$$

are the MLEs of location parameters of the model. If the covariance matrices are unknown, give the MLEs of covariances and hence the MLEs of location parameters.

8.2 Consider again model (8.1) with the known covariance matrices. Show that the MLEs of location parameters are also unbiased estimates.

8.3 Consider the repeated measures problem with two treatment groups. Assume model (8.1) with unequal group variances. Consider the left-sided hypotheses of the form

$$H_0 : \theta_{g_1} - \theta_{g_2} \leq \theta_0.$$

(a) Show that the generalized p-value for testing the hypothesis is given by

$$p = 1 - EG_{n_{g_1}+n_{g_2}-2}\left[\frac{(\overline{y}_{g1} - \overline{y}_{g2} - \theta_0)\sqrt{n_{g_1} + n_{g_2} - 2}}{\sqrt{\frac{1}{T}\left(\frac{s_{g1}}{Bn_{g1}} + \frac{s_{g2}}{(1-B)n_{g2}}\right)}}\right],$$

where $G_{n_{g_1}+n_{g_2}-2}$ is the cumulative distribution function of the Student's t distribution with $n_{g_1} + n_{g_2} - 2$ degrees of freedom and the expectation is with respect to the beta random variable

$$B = \frac{V_{g_1}}{V_{g_1} + V_{g_2}} \sim \text{Beta}\left(\frac{n_{g_1} - 1}{2}, \frac{n_{g_2} - 1}{2}\right).$$

(b) Show that the test is unbiased.
(c) What is the generalized test for testing the right-sided hypotheses of the form

$$H_0 : \theta_{g_1} - \theta_{g_2} \geq \theta_0?$$

(d) What is the equal-tail generalized test for testing the point null hypotheses of the form

$$H_0 : \theta_{g_1} - \theta_{g_2} = \theta_0?$$

8.4 Consider again the multiple comparison problem in Exercise 8.3.

(a) Construct a generalized pivotal quantity for interval estimation of $\theta = \theta_{g_1} - \theta_{g_2}$.
(b) Show that the left-sided $100\gamma\%$ generalized confidence for θ is of the form

$$\theta \leq (\overline{y}_{g1} - \overline{y}_{g2}) + k_\gamma(s_{g_1}, s_{g2}),$$

where $k_\gamma = k_\gamma(s_{g_1}, s_{g2})$ is chosen such that

$$EG_{n_{g_1}+n_{g_2}-2}\left[\frac{k_\gamma\sqrt{n_{g_1}+n_{g_2}-2}}{\sqrt{\frac{1}{T}(\frac{s_{g1}}{Bn_{g1}} + \frac{s_{g2}}{(1-B)n_{g2}})}}\right] = \gamma.$$

8.5 Consider the two-factor repeated measures model under heteroscedasticity. By constructing an extreme region similar to the one used in (8.22) or otherwise derive generalized p-values for comparing two occasion means β_1 and β_2. Deduce generalized confidence intervals for $\beta = \beta_1 - \beta_2$. Describe how one can perform multiple comparisons of occasion means.

8.6 Consider again the two-factor repeated measures model under heteroscedasticity. By taking an approach similar to the one in Exercise 8.5, obtain generalized p-values and generalized confidence intervals for multiple comparison of interaction effects.

8.7 Show that the region of the sample space defined by (8.39) is an extreme region appropriate for defining generalized p-values. Deduce the form of two-sided generalized confidence intervals based on the p-value.

8.8 Show that the generalized p-value for ANCOVA given by (8.51) can also be expressed as a generalized F-test in which the expectation is taken with respect to a set of beta random variables.

8.9 Assuming model (8.43), establish procedures for testing the occasion effects and the interaction between the occasion and treatment effects.

8.10 Consider the data reported in Table 7.4. Carry out a repeated measures ANOVA without the assumption of equal group variances. Compare your results with the results discussed in Chapter 7 under the assumption of equal variances.

8.11 The table below presents a data set reported by Ho, Weerahandi, and Hung (2002) relating to a pharmaceutical study on drug for treating male erectile dysfunction. The objective of the study was to determine the efficacy and safety of active treatment at dose levels 0.5 mg, 1.5 mg, and 5 mg compared to placebo.

Bi-weekly response data

	Time period:	1	2	3	4	5
Group	Subject					
1	1	29.90	29.22	30.91	31.35	29.27
1	2	29.75	30.90	29.11	31.76	29.01
1	3	29.16	32.64	31.68	30.19	30.30
1	4	29.33	27.54	29.16	33.77	31.21
1	5	29.28	29.93	31.35	29.03	30.64
2	6	27.06	32.16	31.88	35.33	38.35
2	7	32.08	38.20	21.75	29.53	33.51
2	8	20.74	23.14	34.34	33.63	29.96
2	9	30.53	31.57	35.36	33.04	24.58
2	10	31.27	21.92	41.40	38.48	22.38
3	11	30.99	30.85	32.49	33.13	33.58
3	12	32.68	30.68	30.30	33.11	32.76
3	13	30.99	33.33	31.09	31.78	31.87
3	14	31.83	31.58	33.62	34.79	33.89
3	15	34.13	31.09	33.44	31.52	29.31
4	16	24.77	27.53	28.88	34.23	35.21
4	17	35.62	32.55	29.92	26.54	28.22
4	18	36.16	31.28	33.63	40.04	33.28
4	19	33.84	29.90	26.45	30.10	30.76
4	20	28.01	30.62	30.91	25.20	28.53

(a) Perform the classical ANOVA under the assumption of equal group variances.
(b) Test the equality of main effects and interactions without the assumption of equal variances, and discuss your findings.
(c) Carry out all pairwise simultaneous comparisons on treatment means and discuss your findings.
(d) Construct 99% confidence intervals for each pair of treatment means.

8.12 Consider again the data set reported in Exercise 7.10. Carry out repeated measures ANOVA without the assumption of equal group variances. Compare your results with the results obtained under the equal variances assumption. Construct 95% simultaneous confidence intervals for each pair of mean effects of the treatment groups.

8.13 Consider the data from Table 7.4, an example on repeated measures reported by Crowder and Hand (1990). Without assuming equal group variances, carry out an RM ANCOVA to compare the diet supplements in a regression setting in which the vector of response means prior to administering the diets is included as a covariate in the regression. Also test the effects of time and the interaction effects.