Chapter 13
Multivariate Analysis of Variation

13.1. Introduction

We will employ the same notations as in the previous chapters. Lower-case letters $x, y, \ldots$ will denote real scalar variables, whether mathematical or random. Capital letters $X, Y, \ldots$ will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed on top of letters such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$ to denote variables in the complex domain. Constant matrices will for instance be denoted by A, B, C. A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. The determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ and, in the complex case, the absolute value or modulus of the determinant of A will be denoted as $|\det(A)|$. When matrices are square, their order will be taken as $p \times p$, unless specified otherwise. When A is a full rank matrix in the complex domain, then AA^* is Hermitian positive definite where an asterisk designates the complex conjugate transpose of a matrix. Additionally, $\mathrm{d}X$ will indicate the wedge product of all the distinct differentials of the elements of the matrix X. Thus, letting the $p \times q$ matrix $X = (x_{ij})$ where the x_{ij}'s are distinct real scalar variables, $\mathrm{d}X = \wedge_{i=1}^{p} \wedge_{j=1}^{q} \mathrm{d}x_{ij}$. For the complex matrix $\tilde{X} = X_1 + iX_2$, $i = \sqrt{(-1)}$, where X_1 and X_2 are real, $\mathrm{d}\tilde{X} = \mathrm{d}X_1 \wedge \mathrm{d}X_2$.

In this chapter, we only consider analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) problems involving real populations. Even though all the steps involved in the following discussion focusing on the real variable case can readily be extended to the complex domain, it does not appear that a parallel development of analysis of variance methodologies in the complex domain has yet been considered. In order to elucidate the various steps in the procedures, we will first review the univariate case. For a detailed exposition of the analysis of variance technique in the scalar variable case, the reader may refer Mathai and Haubold (2017). We will consider the cases of one-way classification or completely randomized design as well as two-way classification with-

out and with interaction or randomized block design. With this groundwork in place, the derivations of the results in the multivariate setting ought to prove easier to follow.

In the early nineteenth century, Gauss and Laplace utilized methodologies that may be regarded as forerunners to ANOVA in their analyses of astronomical data. However, this technique came to full fruition in Ronald Fisher's classic book titled *"Statistical Methods for Research Workers"*, which was initially published in 1925. The principle behind ANOVA consists of partitioning the total variation present in the data into variations attributable to different sources. It is actually the total variation that is split rather than the total variance, the latter being a fraction of the former. Accordingly, the procedure could be more appropriately referred to as "analysis of variation". As has already been mentioned, we will initially consider the one-way classification model, which will then be extended to the multivariate situation.

Let us first focus on an experimental design called a completely randomized experiment. In this setting, the subject matter was originally developed for agricultural experiments, which influenced its terminology. For example, the basic experimental unit is referred to as a "plot", which is a piece of land in an agricultural context. When an experiment is performed on human beings, a plot translates into an individual. If the experiment is carried out on some machinery, then a machine corresponds to a plot. In a completely randomized experiment, a set of $n_1 + n_2 + \cdots + n_k$ plots, which are homogeneous with respect to all factors of variation, are selected. Then, k treatments are applied at random to these plots, the first treatment to n_1 plots, the second treatment to n_2 plots, up to the k-th treatment being applied to n_k plots. For instance, if the effects of k different fertilizers on the yield of a certain crop are to be studied, then the treatments consist of these k fertilizers, the first treatment meaning one of the fertilizers, the second treatment, another one and so on, with the k-th treatment corresponding to the last fertilizer. If the experiment involves studying the yield of corn among k different varieties of corn, then a treatment coincides with a particular variety. If an experiment consists of comparing k teaching methods, then a treatment refers to a method of teaching and a plot corresponds to a student. When an experiment compares the effect of k different medications in curing a certain ailment, then a treatment is a medication, and so on. If the treatments are denoted by $t_1, \ldots, t_k$, then treatment t_j is applied at random to n_j homogeneous plots or n_j homogeneous plots are selected at random and treatment t_j is applied to them, for $j = 1, \ldots, k$. Random assignment is done to avoid possible biases or the influence of confounding factors, if any. Then, observations measuring the effect of these treatments on the experimental units are made. For example, in the case of various methods of teaching, the observation x_{ij} could be the final grade obtained by the j-th student who was subjected to the i-th teaching method. In the case of comparing k different varieties of corn, the observation x_{ij} could consist of the yield of corn observed at harvest time in the j-th plot which received the i-th variety of corn. Thus, in this instance, i stands for the treatment number and j represents the serial

number of the plot receiving the i-th treatment, x_{ij} being the final observation. Then, the corresponding linear additive fixed effect model is the following:

$$x_{ij} = \mu + \alpha_i + e_{ij}, \ j = 1, \ldots, n_i, \ i = 1, \ldots, k, \tag{13.1.1}$$

where μ is a general effect, α_i is the deviation from the general effect due to treatment t_i and e_{ij} is the random component, which includes the sum total contributions originating from unknown or uncontrolled factors. When the experiment is designed, the plots are selected so that they be homogeneous with respect to all possible factors of variation. The general effect μ can be interpreted as the grand average or the expected value of x_{ij} when α_i is not present or treatment t_i is not applied or has no effect. The simplest assumption that we will make is that $E[e_{ij}] = 0$ for all i and j and $\text{Var}(e_{ij}) = \sigma^2 > 0$ for all i and j and for some positive quantity σ^2, where $E[\cdot]$ denotes the expected value of $[\cdot]$. It is further assumed that $\mu, \alpha_1, \ldots, \alpha_k$ are all unknown constants. When $\alpha_1, \ldots, \alpha_k$ are assumed to be random variables, model (13.1.1) is referred to as a"random effect model". In the following discussion, we will solely consider fixed effect models. The first step consists of estimating the unknown quantities from the data. Since no distribution is assumed on the e_{ij}'s, and thereby on the x_{ij}'s, we will employ the method of least squares for estimating the parameters. In that case, one has to minimize the error sum of squares which is

$$\sum_{ij} e_{ij}^2 = \sum_{ij} [x_{ij} - \mu - \alpha_i]^2.$$

Applying calculus principles, we equate the partial derivatives of $\sum_{ij} e_{ij}^2$ with respect to μ to zero and then, equate the partial derivatives of $\sum_{ij} e_{ij}^2$ with respect to $\alpha_1, \ldots, \alpha_k$ to zero and solve these equations. A convenient notation in this area is to represent a summation by a dot. As an example, if the subscript j is summed up, it is replaced by a dot, so that $\sum_j x_{ij} \equiv x_{i.}$; similarly, $\sum_{ij} x_{ij} \equiv x_{..}$. Thus,

$$\frac{\partial}{\partial \mu} \Big[\sum_{ij} e_{ij}^2 \Big] = 0 \Rightarrow -2 \sum_{ij} [x_{ij} - \mu - \alpha_i] = 0 \Rightarrow \sum_i \Big(\sum_j [x_{ij} - \mu - \alpha_i] \Big) = 0$$

that is,

$$\sum_i [x_{i.} - n_i \mu - n_i \alpha_i] = 0 \Rightarrow x_{..} - n_{.} \mu - \sum_{i=1}^{k} n_i \alpha_i = 0,$$

and since we have taken α_i as a deviation from the general effect due to treatment t_i, we can let $\sum_i n_i \alpha_i = 0$ without any loss of generality. Then, $x_{..}/n_{.}$ is an estimate of μ, and denoting estimates/estimators by a hat, we write $\hat{\mu} = x_{..}/n_{.}$. Now, note that for example

α_1 appears in the terms $(x_{11} - \mu - \alpha_1)^2 + \cdots + (x_{1n_1} - \mu - \alpha_1)^2 = \sum_j (x_{1j} - \mu - \alpha_1)^2$ but does not appear in the other terms in the error sum of squares. Accordingly, for a specific i,

$$\frac{\partial}{\partial \alpha_i} \left[\sum_{ij} e_{ij}^2 \right] = 0 \Rightarrow \sum_j [x_{ij} - \mu - \alpha_i] = 0 \Rightarrow x_{i.} - n_i \hat{\mu} - n_i \hat{\alpha}_i = 0,$$

that is, $\hat{\alpha}_i = \frac{x_{i.}}{n_i} - \hat{\mu}$. Thus,

$$\hat{\mu} = \frac{1}{n_.} x_{..} \text{ and } \hat{\alpha}_i = \frac{1}{n_i} x_{i.} - \hat{\mu}. \tag{13.1.2}$$

The least squares minimum is obtained by substituting the least squares estimates of μ and α_i, $i = 1, \ldots, k$, in the error sum of squares. Denoting the least squares minimum by s^2,

$$s^2 = \sum_{ij} (x_{ij} - \hat{\mu} - \hat{\alpha}_i)^2 = \sum_{ij} \left[x_{ij} - \frac{x_{..}}{n_.} - \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n_.} \right) \right]^2$$

$$= \sum_{ij} \left[x_{ij} - \frac{x_{i.}}{n_i} \right]^2 = \sum_{ij} \left[x_{ij} - \frac{x_{..}}{n_.} \right]^2 - \sum_{ij} \left[\frac{x_{i.}}{n_i} - \frac{x_{..}}{n_.} \right]^2. \tag{13.1.3}$$

When the square is expanded, the middle term will become $-2 \sum_{ij} (\frac{x_{i.}}{n_i} - \frac{x_{..}}{n_.})^2$, thus yielding the expression given in (13.1.3). As well, we have the following identity:

$$\sum_{ij} \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n_.} \right)^2 = \sum_{i=1}^{k} n_i \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n_.} \right)^2 = \sum_i \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{n_.}.$$

Now, let us consider the hypothesis $H_o : \alpha_1 = \alpha_2 = \cdots = \alpha_k$, which is equivalent to the hypothesis $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$ since, by assumption, $\sum_i n_i \alpha_i = 0$. Proceeding as before, the least squares minimum, under the null hypothesis H_o, denoted by s_0^2, is the following:

$$s_0^2 = \sum_{ij} \left(x_{ij} - \frac{x_{..}}{n_.} \right)^2$$

and hence the sum of squares due to the hypothesis or due to the presence of the α_j's, is given by $s_0^2 - s^2 = \sum_{ij} (\frac{x_{i.}}{n_i} - \frac{x_{..}}{n_.})^2$. Thus, the total variation is partitioned as follows:

$$s_0^2 = [s_0^2 - s^2] + [s^2]$$

$$\sum_{ij} \left(x_{ij} - \frac{x_{..}}{n_.} \right)^2 = \left[\sum_i n_i \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n_.} \right)^2 \right] + \left[\sum_{ij} \left(x_{ij} - \frac{x_{i.}}{n_i} \right)^2 \right], \text{ that is,}$$

Total variation (s_0^2)=variation due to the α_j's $(s_0^2 - s^2)$+the residual variation (s^2),

which is the analysis of variation principle. If $e_{ij} \overset{iid}{\sim} N_1(0, \sigma^2)$ for all i and j where $\sigma^2 > 0$ is a constant, it follows from the chisquaredness and independence of quadratic forms, as discussed in Chaps. 2 and 3, that $\frac{s_0^2}{\sigma^2} \sim \chi^2_{n.-1}$, a real chisquare variable having $n. - 1$ degrees of freedom, $\frac{[s_0^2 - s^2]}{\sigma^2} \sim \chi^2_{k-1}$ under the hypothesis H_o and $\frac{s^2}{\sigma^2} \sim \chi^2_{n.-k}$, where the sum of squares due to the α_j's, namely $s_0^2 - s^2$, and the residual sum of squares, namely s^2, are independently distributed under the hypothesis. Usually, these findings are put into a tabular form known as the analysis of variation table or ANOVA table. The usual format is as follows:

ANOVA Table for the One-Way Classification

Variation due to (1)	df (2)	SS (3)	MS (3)/(2)
treatments	$k - 1$	$\sum_i n_i (\frac{x_{i.}}{n_i} - \frac{x_{..}}{n.})^2$	$(s_0^2 - s^2)/(k - 1)$
residuals	$n. - k$	$\sum_{ij} (x_{ij} - \frac{x_{i.}}{n_i})^2$	$s^2/(n. - k)$
total	$n. - 1$	$\sum_{ij} (x_{ij} - \frac{x_{..}}{n.})^2$	

where df denotes the number of degrees of freedom, SS means sum of squares and MS stands for mean squares or the average of the squares. There is usually a last column which contains the F-ratio, that is, the ratio of the treatments MS to the residuals MS, and enables one to determine whether to reject the null hypothesis, in which case the test statistic is said to be "significant", or not to reject the null hypothesis, when the test statistic is "not significant". Further details on the real scalar variable case are available from Mathai and Haubold (2017).

In light of this brief review of the scalar variable case of one-way classification data or univariate data secured from a completely randomized design, the concepts will now be extended to the multivariate setting.

13.2. Multivariate Case of One-Way Classification Data Analysis

Extension of the results to the multivariate case is parallel to the scalar variable case. Consider a model of the type

$$X_{ij} = M + A_i + E_{ij}, \ j = 1, \ldots, n_i, \ i = 1, \ldots, k, \tag{13.2.1}$$

with X_{ij}, M, A_i and E_{ij} all being $p \times 1$ real vectors where X_{ij} denotes the j-th observation vector in the i-th group or the observed vectors obtained from the n_i plots receiving

the i-th vector of treatments, M is a general effect vector, A_i is a vector of deviations from M due to the i-th treatment vector so that $\sum_i n_i A_i = O$ since we are taking deviations from the general effect M, and E_{ij} is a vector of random components assumed to be normally distributed as follows: $E_{ij} \overset{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$, for all i and j where Σ is a positive definite covariance matrix, that is,

$$\text{Cov}(E_{ij}) = E[(E_{ij} - O)(E_{ij} - O)'] = E[E_{ij} E_{ij}'] = \Sigma > O \text{ for all } i \text{ and } j,$$

where $E[\,\cdot\,]$ denotes the expected value operator. This normality assumption will be needed for testing hypotheses and developing certain distributional aspects. However, the multivariate analysis of variation can be set up without having to resort to any distributional assumption. In the real scalar variable case, we minimized the sum of the squares of the errors since the variations only involved single scalar variables. In the vector case, if we take the sum of squares of the elements in E_{ij}, that is, $E_{ij}' E_{ij}$ and its sum over all i and j, then we are only considering the variations in the individual elements of E_{ij}'s; however, in the vector case, there is joint variation among the elements of the vector and that is also to be taken into account. Hence, we should be considering all squared terms and cross product terms or the whole matrix of squared and cross product terms. This is given by $E_{ij} E_{ij}'$ and so, we should consider this matrix and carry out some type of minimization. Consider

$$\sum_{ij} E_{ij} E_{ij}' = \sum_{ij} [X_{ij} - M - A_i][X_{ij} - M - A_i]'. \tag{13.2.2}$$

For obtaining estimates of M and A_i, $i = 1, \ldots, k$, we will minimize the trace of $\sum_{ij} E_{ij} E_{ij}'$ as a criterion. There are terms of the type $[X_{ij} - M - A_i]'[X_{ij} - M - A_i]$ in this trace. Thus,

$$\frac{\partial}{\partial M}\left[\text{tr}\left(\sum_{ij} E_{ij} E_{ij}'\right)\right] = \sum_{ij} \frac{\partial}{\partial M}[X_{ij} - M - A_i]'[X_{ij} - M - A_i] = O$$

$$\Rightarrow \sum_{ij} X_{ij} - n_. M - \sum_i n_i A_i = O \Rightarrow \hat{M} = \frac{1}{n_.} X_{..},$$

noting that we assumed that $\sum_i n_i A_i = O$. Now, on differentiating the trace of $E_{ij} E_{ij}'$ with respect to A_i for a specific i, we have

$$\frac{\partial}{\partial A_i}\text{tr}\left[\sum_{ij} E_{ij} E_{ij}'\right] = \frac{\partial}{\partial A_i} \sum_{ij} [X_{ij} - M - A_i]'[X_{ij} - M - A_i] = O$$

$$\Rightarrow \sum_j [X_{ij} - M - A_i] = O \Rightarrow \hat{A}_i = \frac{1}{n_i} X_{i.} - \hat{M} = \frac{1}{n_i} X_{i.} - \frac{1}{n_.} X_{..}.$$

Observe that there is only one critical vector for $\hat{M}$ and for $\hat{A}_i$, $i = 1, \ldots, k$. Accordingly, the critical point will either correspond to a minimum or a maximum of the trace. But for arbitrary M and A_i, the maximum occurs at plus infinity and hence, the critical point $(\hat{M}, \hat{A}_i, \ i = 1, \ldots, k)$ corresponds to a minimum. Once evaluated at these estimates, the sum of squares and cross products matrix, denoted by S, is the following:

$$S = \sum_{ij}[X_{ij} - \hat{M} - \hat{A}_i][X_{ij} - \hat{M} - \hat{A}_i]' = \sum_{ij}\left[X_{ij} - \frac{1}{n_i}X_{i.}\right]\left[X_{ij} - \frac{1}{n_i}X_{i.}\right]'$$

$$= \sum_{ij}\left[X_{ij} - \frac{1}{n_.}X_{..}\right]\left[X_{ij} - \frac{1}{n_.}X_{..}\right]' - \sum_{i}n_i\left[\frac{1}{n_i}X_{i.} - \frac{1}{n_.}X_{..}\right]\left[\frac{1}{n_i}X_{i.} - \frac{1}{n_.}X_{..}\right]'$$

$$(13.2.3)$$

Note that as in the scalar case, the middle terms and the last term will combine into the second term above. Now, let us impose the hypothesis $H_o : A_1 = A_2 = \cdots = A_k = O$. Note that equality of the A_j's will automatically imply that each one is null because the weighted sum is null as per our initial assumption in the model (13.2.1). Under this hypothesis, the model will be $X_{ij} = M + E_{ij}$, and then proceeding as in the univariate case, we end up with the following sum of squares and cross products matrix, denoted by S_0:

$$S_0 = \sum_{ij}\left[X_{ij} - \frac{1}{n_.}X_{..}\right]\left[X_{ij} - \frac{1}{n_.}X_{..}\right]', \tag{13.2.4}$$

so that the sum of squares and cross products matrix due to the A_i's is the difference

$$S_0 - S = \sum_{ij}\left[\frac{1}{n_i}X_{i.} - \frac{1}{n_.}X_{..}\right]\left[\frac{1}{n_i}X_{i.} - \frac{1}{n_.}X_{..}\right]'. \tag{13.2.5}$$

Thus, the following partitioning of the total variation in the multivariate data:

$$S_0 = [S_0 - S] + S$$

$$\text{Total variation} = [\text{Variation due to the } A_i\text{'s}] + [\text{Residual variation}]$$

$$\sum_{ij}\left[X_{ij} - \frac{1}{n_.}X_{..}\right]\left[X_{ij} - \frac{1}{n_.}X_{..}\right]' = \sum_{ij}\left[\frac{1}{n_i}X_{i.} - \frac{1}{n_.}X_{..}\right]\left[\frac{1}{n_i}X_{i.} - \frac{1}{n_.}X_{..}\right]'$$

$$+ \sum_{ij}\left[X_{ij} - \frac{1}{n_i}X_{i.}\right]\left[X_{ij} - \frac{1}{n_i}X_{i.}\right]'.$$

Under the normality assumption for the random component $E_{ij} \overset{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$, we have the following properties, which follow from results derived in Chap. 5, the

notation $W_p(\nu, \Sigma)$ standing for a Wishart distribution having ν degrees of freedom and parameter matrix Σ:

$$\text{Total variation} = S_0 = \sum_{ij} \left[X_{ij} - \frac{1}{n_.} X_{..} \right] \left[X_{ij} - \frac{1}{n_.} X_{..} \right]',$$

$$S_0 \sim W_p(n_. - 1, \Sigma);$$

$$\text{Variation due to the } A_i\text{'s} = S_0 - S = \sum_{ij} \left[\frac{1}{n_i} X_{i.} - \frac{1}{n_.} X_{..} \right] \left[\frac{1}{n_i} X_{i.} - \frac{1}{n_.} X_{..} \right]',$$

$$S_0 - S \sim W_p(k - 1, \Sigma) \text{ under the hypothesis } A_1 = A_2 = \cdots = A_k = O;$$

$$\text{Residual variation} = S = \sum_{ij} \left[X_{ij} - \frac{1}{n_i} X_{i.} \right] \left[X_{ij} - \frac{1}{n_i} X_{i.} \right]',$$

$$S \sim W_p(n_. - k, \Sigma).$$

We can summarize these findings in a tabular form known as the multivariate analysis of variation table or MANOVA table, where df means degrees of freedom in the corresponding Wishart distribution, and SSP represents the sum of squares and cross products matrix.

Multivariate Analysis of Variation (MANOVA) Table

Variation due to	df	SSP
treatments	$k - 1$	$\sum_{ij} [\frac{1}{n_i} X_{i.} - \frac{1}{n_.} X_{..}][\frac{1}{n_i} X_{i.} - \frac{1}{n_.} X_{..}]'$
residuals	$n_. - k$	$\sum_{ij} [X_{ij} - \frac{1}{n_i} X_{i.}][X_{ij} - \frac{1}{n_i} X_{i.}]'$
total	$n_. - 1$	$\sum_{ij} [X_{ij} - \frac{1}{n_.} X_{..}][X_{ij} - \frac{1}{n_.} X_{..}]'$

13.2.1. Some properties

The sample values from the i-th sample or the i-th group or the plots receiving the i-th treatment are $X_{i1}, X_{i2}, \ldots, X_{in_i}$. In this case, the average is $\sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} = \frac{X_{i.}}{n_i}$ and the i-th sample sum of squares and products matrix is

$$S_i = \sum_{j=1}^{n_i} \left(X_{ij} - \frac{X_{i.}}{n_i} \right) \left(X_{ij} - \frac{X_{i.}}{n_i} \right)'.$$

As well, it follows from Chap. 5 that $S_i \sim W_p(n_i - 1, \Sigma)$ when $E_{ij} \overset{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$. Then, the residual sum of squares and products matrix can be written as follows, denoting it by the matrix V:

$$V = \sum_{ij} \left(X_{ij} - \frac{X_{i.}}{n_i} \right)\left(X_{ij} - \frac{X_{i.}}{n_i} \right)' = \sum_{i=1}^{k} \left[\sum_{j=1}^{n_i} \left(X_{ij} - \frac{X_{i.}}{n_i} \right)\left(X_{ij} - \frac{X_{i.}}{n_i} \right)' \right]$$

$$= \sum_{i=1}^{k} S_i = S_1 + S_2 + \cdots + S_k \tag{13.2.6}$$

where $S_i \sim W_p(n_i - 1, \Sigma)$, $i = 1, \ldots, k$, and the S_i's are independently distributed since the sample values from the k groups are independently distributed among themselves (within the group) and between groups. Hence, $S \sim W_p(\nu, \Sigma)$, $\nu = (n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) = n_. - k$. Note that $\bar{X}_i = \frac{X_{i.}}{n_i}$ has $\text{Cov}(\bar{X}_i) = \frac{1}{n_i}\Sigma$, so that $\sqrt{n_i}(\bar{X}_i - \bar{X})$ are iid $N_p(O, \Sigma)$ where $\bar{X} = X_{..}/n_.$. Then, the sum of squares and products matrix due to the treatments or due to the A_i's is the following, denoting it by U:

$$U = \sum_{i=1}^{k} n_i \left[\frac{X_{i.}}{n_i} - \frac{X_{..}}{n_.} \right]\left[\frac{X_{i.}}{n_i} - \frac{X_{..}}{n_.} \right]' \sim W_p(k - 1, \Sigma) \tag{13.2.7}$$

under the null hypothesis; when the hypothesis is violated, it is a noncentral Wishart distribution. Further, the sum of squares and products matrix due to the treatments and the residual sum of squares and products matrix are independently distributed. Thus, by comparing U and V, we should be able to reach a decision regarding the hypothesis. One procedure that is followed is to take the determinants of U and V and compare them. This does not have much of a basis and determinants should not be called "generalized variance" as previously explained since the basic condition of a norm is violated by the determinant. The basis for comparing determinants will become clear from the point of view of testing hypotheses by applying the likelihood ratio criterion, which is discussed next.

13.3. The Likelihood Ratio Criterion

Let $E_{ij} \overset{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$, and suppose that we have simple random samples of sizes $n_1, \ldots, n_k$ from the k groups relating to the k treatments. Then, the likelihood function, denoted by L, is the following:

$$L = \prod_{ij} \frac{e^{-\frac{1}{2}(X_{ij} - M - A_i)'\Sigma^{-1}(X_{ij} - M - A_i)}}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}}$$

$$= \frac{e^{-\frac{1}{2}\sum_{ij}(X_{ij} - M - A_i)'\Sigma^{-1}(X_{ij} - M - A_i)}}{(2\pi)^{\frac{pn_.}{2}}|\Sigma|^{\frac{n_.}{2}}}. \tag{13.3.1}$$

The maximum likelihood estimators/estimates (MLE's) of M is $\hat{M} = \frac{X_{..}}{n_{.}} = \bar{X}$ and that of A_i is $\hat{A}_i = \frac{X_{i.}}{n_i} - \hat{M}$. With a view to obtaining the MLE of Σ, we first note that the exponent is a real scalar quantity which is thus equal to its trace, so that we can express the exponent as follows, after substituting the MLE's of M and A_i:

$$-\frac{1}{2}\sum_{ij}[X_{ij} - \hat{M} - \hat{A}_i]'\Sigma^{-1}[X_{ij} - \hat{M} - \hat{A}_i]$$

$$= -\frac{1}{2}\sum_{ij}\text{tr}\left(\left[X_{ij} - \frac{X_{i.}}{n_i}\right]'\Sigma^{-1}\left[X_{ij} - \frac{X_{i.}}{n_i}\right]\right)$$

$$= -\frac{1}{2}\sum_{ij}\text{tr}\left(\Sigma^{-1}\left[X_{ij} - \frac{X_{i.}}{n_i}\right]\left[X_{ij} - \frac{X_{i.}}{n_i}\right]'\right).$$

Now, following through the estimation procedure of the MLE included in Chap. 3, we obtain the MLE of Σ as

$$\hat{\Sigma} = \frac{1}{n_{.}}\sum_{ij}\left(X_{ij} - \frac{X_{i.}}{n_i}\right)\left(X_{ij} - \frac{X_{i.}}{n_i}\right)'. \tag{13.3.2}$$

After substituting $\hat{M}$, $\hat{A}_i$ and $\hat{\Sigma}$, the exponent in the likelihood ratio criterion λ becomes $-\frac{1}{2}n_{.}\text{tr}(I_p) = -\frac{1}{2}n_{.}p$. Hence, the maximum value of the likelihood function L under the general model becomes

$$\max L = \frac{e^{-\frac{1}{2}(n_{.}p)}n_{.}^{\frac{n_{.}p}{2}}}{(2\pi)^{\frac{n_{.}p}{2}}|\sum_{ij}(X_{ij} - \frac{X_{i.}}{n_i})(X_{ij} - \frac{X_{i.}}{n_i})'|}. \tag{13.3.3}$$

Under the hypothesis $H_o : A_1 = A_2 = \cdots = A_p = O$, the model is $X_{ij} = M + E_{ij}$ and the MLE of M under H_o is still $\frac{1}{n_{.}}X_{..}$ and $\hat{\Sigma}$ under H_o is $\frac{1}{n_{.}}\sum_{ij}(X_{ij} - \frac{1}{n_{.}}X_{..})(X_{ij} - \frac{1}{n_{.}}X_{..})'$, so that $\max L$ under H_o, denoted by $\max L_o$, is

$$\max L_o = \frac{e^{-\frac{1}{2}n_{.}p}n_{.}^{\frac{n_{.}p}{2}}}{(2\pi)^{\frac{n_{.}p}{2}}|\sum_{ij}(X_{ij} - \frac{1}{n_{.}}X_{..})(X_{ij} - \frac{1}{n_{.}}X_{..})'|^{\frac{n_{.}}{2}}}. \tag{13.3.4}$$

Therefore, the λ-criterion is the following:

$$\lambda = \frac{\max L_o}{\max L} = \frac{|\sum_{ij}(X_{ij} - \frac{X_{i.}}{n_i})(X_{ij} - \frac{X_{i.}}{n_i})'|^{\frac{n_{.}}{2}}}{|\sum_{ij}(X_{ij} - \frac{X_{..}}{n_{.}})(X_{ij} - \frac{X_{..}}{n_{.}})'|^{\frac{n_{.}}{2}}}$$

$$= \frac{|V|^{\frac{n_{.}}{2}}}{|U + V|^{\frac{n_{.}}{2}}} \tag{13.3.5}$$

where

$$U = \sum_{ij} \Big(\frac{X_{i.}}{n_i} - \frac{X_{..}}{n_{.}}\Big)\Big(\frac{X_{i.}}{n_i} - \frac{X_{..}}{n_{.}}\Big)', \ V = \sum_{ij} \Big(X_{ij} - \frac{X_{i.}}{n_i}\Big)\Big(X_{ij} - \frac{X_{i.}}{n_i}\Big)'$$

and $U \sim W_p(k-1, \Sigma)$ under H_o is the sum of squares and cross products matrix due to the A_i's and $V \sim W_p(n_{.} - k, \Sigma)$ is the residual sum of squares and cross products matrix. It has already been shown that U and V are independently distributed. Then $W_1 = (U+V)^{-\frac{1}{2}}V(U+V)^{-\frac{1}{2}}$, with the determinant $\frac{|V|}{|U+V|}$, is a real matrix-variate type-1 beta with parameters $(\frac{n_{.}-k}{2}, \frac{k-1}{2})$, as defined in Chap. 5, and $W_2 = V^{-\frac{1}{2}}UV^{-\frac{1}{2}}$ is a real matrix-variate type-2 beta with the parameters $(\frac{k-1}{2}, \frac{n_{.}-k}{2})$. Moreover, $Y_1 = I - W_1 = (U+V)^{-\frac{1}{2}}U(U+V)^{-\frac{1}{2}}$ with $\frac{|U|}{|U+V|}$ is a real matrix-variate type-1 beta random variables with parameters $(\frac{k-1}{2}, \frac{n_{.}-k}{2})$. Given the properties of independent real matrix-variate gamma random variables, we have seen in Chap. 5 that W_1 and $Y_2 = U + V$ are independently distributed. Similarly, $Y_1 = I - W_1$ and Y_2 are independently distributed. Further, $W_2^{-1} = V^{\frac{1}{2}}U^{-1}V^{\frac{1}{2}}$ is a real matrix-variate type-2 beta random variable with the parameters $(\frac{n_{.}-k}{2}, \frac{k-1}{2})$. Observe that

$$|W_1| = \frac{|V|}{|U+V|} = \frac{1}{|V^{-\frac{1}{2}}UV^{-\frac{1}{2}} + I|} = \frac{1}{|W_2 + I|}, \ W_1 = (I + W_2)^{-1}.$$

A one-to-one function of λ is

$$w = \lambda^{\frac{2}{n_{.}}} = \frac{|V|}{|U+V|} = |W_1|. \tag{13.3.6}$$

13.3.1. Arbitrary moments of the likelihood ratio criterion

For an arbitrary h, the h-th moment of w as well as that of λ can be obtained from the normalizing constant of a real matrix-variate type-1 beta density with the parameters $(\frac{n_{.}-k}{2}, \frac{k-1}{2})$. That is,

$$E[w^h] = \frac{\Gamma_p(\frac{n_{.}-k}{2} + h)}{\Gamma_p(\frac{n_{.}-k}{2})} \frac{\Gamma_p(\frac{n_{.}-1}{2})}{\Gamma_p(\frac{n_{.}-1}{2} + h)}, \ (n_{.} - k) + (k - 1) = n_{.} - 1,$$

$$= \Bigg\{\prod_{j=1}^{p} \frac{\Gamma(\frac{n_{.}-1}{2} - \frac{j-1}{2})}{\Gamma(\frac{n_{.}-k}{2} - \frac{j-1}{2})}\Bigg\} \Bigg\{\prod_{j=1}^{p} \frac{\Gamma(\frac{n_{.}-k}{2} - \frac{j-1}{2} + h)}{\Gamma(\frac{n_{.}-1}{2} - \frac{j-1}{2} + h)}\Bigg\}. \tag{13.3.7}$$

As $E[\lambda^h] = E[w^{\frac{n}{2}}]^h = E[w^{(\frac{n}{2})h}]$, the h-th moment of λ is obtained by replacing h by $(\frac{n.}{2})h$ in (13.3.7). That is,

$$E[\lambda^h] = C_{p,k} \left\{ \prod_{j=1}^{p} \frac{\Gamma(\frac{n.-k}{2} - \frac{j-1}{2} + (\frac{n.}{2})h)}{\Gamma(\frac{n.-1}{2} - \frac{j-1}{2} + (\frac{n.}{2})h)} \right\} \tag{13.3.8}$$

where

$$C_{p.k} = \left\{ \prod_{j=1}^{p} \frac{\Gamma(\frac{n.-1}{2} - \frac{j-1}{2})}{\Gamma(\frac{n.-k}{2} - \frac{j-1}{2})} \right\}.$$

13.3.2. Structural representation of the likelihood ratio criterion

It can readily be seen from (13.3.7) that the h-th moment of w is of the form of the h-th moment of a product of independently distributed real scalar type-1 beta random variables. That is,

$$E[w^h] = E[w_1 w_2 \cdots w_p]^h, \quad w = w_1 w_2 \cdots w_p, \tag{13.3.9}$$

where $w_1, \ldots, w_p$ are independently distributed and w_j is a real scalar type-1 beta random variable with the parameters $(\frac{n.-k}{2} - \frac{j-1}{2}, \frac{k-1}{2})$, $j = 1, \ldots, p$, for $n. - k > p - 1$ and $n. > k + p - 1$. Hence the exact density of w is available by constructing the density of a product of independently distributed real scalar type-1 beta random variables. For special values of p and k, one can obtain the exact densities in the forms of elementary functions. However, for the general case, the exact density corresponding to $E[w^h]$ as specified in (13.3.7) can be expressed in terms of a G-function and, in the case of $E[\lambda^h]$ as given in (13.3.8), the exact density can be represented in terms of an H-function. These representations are as follows, denoting the densities of w and λ as $f_w(w)$ and $f_\lambda(\lambda)$, respectively:

$$f_w(w) = C_{p,k}\, G_{p,p}^{p,0} \left[w \left|_{\frac{n.-k}{2} - \frac{j-1}{2} - 1,\ j=1,\ldots,p}^{\frac{n.-1}{2} - \frac{j-1}{2} - 1,\ j=1,\ldots,p} \right. \right], \quad 0 < w \le 1, \tag{13.3.10}$$

$$f_\lambda(\lambda) = C_{p,k}\, H_{p,p}^{p,0} \left[\lambda \left|_{(\frac{n.-k}{2} - \frac{j-1}{2} - \frac{n.}{2}, \frac{n.}{2}),\ j=1,\ldots,p}^{(\frac{n.-1}{2} - \frac{j-1}{2} - \frac{n.}{2}, \frac{n.}{2}),\ j=1,\ldots,p} \right. \right], \quad 0 < \lambda \le 1, \tag{13.3.11}$$

for $n. > p + k - 1$, $p \ge 1$ and $f_w(w) = 0$, $f_\lambda(\lambda) = 0$, elsewhere. The evaluation of G and H-functions can be carried out with the help of symbolic computing packages such as *Mathematica* and MAPLE. Theoretical considerations, applications and several special cases of the G and H-functions are, for instance, available from Mathai (1993) and Mathai,

Saxena and Haubold (2010). The special cases listed therein can also be utilized to work out the densities for particular cases of (13.3.10) and (13.3.11). Explicit structures of the densities for certain special cases are listed in the next section.

13.3.3. Some special cases

Several particular cases can be worked out by examining the moment expressions in (13.3.7) and (13.3.8). The h-th moment of the $w = \lambda^{\frac{2}{n_.}}$, where λ is the likelihood ratio criterion, is available from (13.3.7) as

$$E[w^h] = C_{p,k} \frac{\Gamma(\frac{n_.-k}{2} + h)\Gamma(\frac{n_.-k}{2} - \frac{1}{2} + h) \cdots \Gamma(\frac{n_.-k}{2} - \frac{p-1}{2} + h)}{\Gamma(\frac{n_.-1}{2} + h)\Gamma(\frac{n_.-1}{2} - \frac{1}{2} + h) \cdots \Gamma(\frac{n_.-1}{2} - \frac{p-1}{2} + h)}. \qquad (i)$$

Case (1): $p = 1$

In this case, from *(i)*,

$$E[w^h] = C_{1,k} \frac{\Gamma(\frac{n_.-k}{2} + h)}{\Gamma(\frac{n_.-1}{2} + h)},$$

which is the h-th moment of a real scalar type-1 beta random variable with the parameters $(\frac{n_.-k}{2}, \frac{k-1}{2})$ and, in this case, w is simply a real scalar type-1 beta random variable with the parameters $(\frac{n_.-k}{2}, \frac{k-1}{2})$. We reject the null hypothesis $H_o : A_1 = A_2 = \cdots = A_k = O$ for small values of the λ-criterion and, accordingly, we reject H_o for small values of w or the hypothesis is rejected when the observed value of $w \le w_\alpha$ where w_α is such that $\int_0^{w_\alpha} f_w(w)\mathrm{d}w = \alpha$ for the preassigned size α of the critical region, $f_w(w)$ denoting the density of w for $p = 1, n_. > k$.

Case (2): $p = 2$

From *(i)*, we have

$$E[w^h] = C_{2,k} \frac{\Gamma(\frac{n_.-k}{2} + h)\Gamma(\frac{n_.-k}{2} - \frac{1}{2} + h)}{\Gamma(\frac{n_.-1}{2} + h)\Gamma(\frac{n_.-1}{2} - \frac{1}{2} + h)}$$

and therefore

$$E[w^{\frac{1}{2}}]^h = C_{2,k} \frac{\Gamma(\frac{n_.-k}{2} + \frac{h}{2})\Gamma(\frac{n_.-k}{2} - \frac{1}{2} + \frac{h}{2})}{\Gamma(\frac{n_.-1}{2} + \frac{h}{2})\Gamma(\frac{n_.-1}{2} - \frac{1}{2} + \frac{h}{2})}. \qquad (ii)$$

The gamma functions in *(ii)* can be combined by making use of a duplication formula for gamma functions, namely,

$$\Gamma(z)\Gamma(z + 1/2) = \pi^{\frac{1}{2}} 2^{1-2z} \Gamma(2z). \qquad (13.3.12)$$

Take $z = \frac{n.-k}{2} - \frac{1}{2} + \frac{h}{2}$ and $z = \frac{n.-1}{2} - \frac{1}{2} + \frac{h}{2}$ in the part containing h and in the constant part wherein $h = 0$, and then apply formula (13.3.12) to obtain

$$E[w^{\frac{1}{2}}]^h = \frac{\Gamma(n. - 2)}{\Gamma(n. - k - 1)} \frac{\Gamma(n. - k - 1 + h)}{\Gamma(n. - 2 + h)},$$

which is, for an arbitrary h, the h-th moment of a real scalar type-1 beta random variable with parameters $(n. - k - 1, k - 1)$ for $n. - k - 1 > 0$, $k > 1$. Thus, $y = w^{\frac{1}{2}}$ is a real scalar type-1 beta random variable with the parameters $(n. - k - 1, k - 1)$. We would then reject H_o for small values of w, that is, for small values of y or when the observed value of $y \leq y_\alpha$ with y_α such that $\int_0^{y_\alpha} f_y(y)dy = \alpha$ for a preassigned probability of type-I error which is the error of rejecting H_o when H_o is true, where $f_y(y)$ is the density of y for $p = 2$ whenever $n. > k + 1$.

Case (3): $k = 2$, $p \geq 1$, $n. > p + 1$

In this case, the h-th moment of w as specified in (13.3.7) is the following:

$$E[w^h] = C_{p,2} \frac{\Gamma(\frac{n.-2}{2} + h)\Gamma(\frac{n.-2}{2} - \frac{1}{2} + h) \cdots \Gamma(\frac{n.-2}{2} - \frac{p-1}{2} + h)}{\Gamma(\frac{n.-1}{2} + h)\Gamma(\frac{n.-1}{2} - \frac{1}{2} + h) \cdots \Gamma(\frac{n.-1}{2} - \frac{p-1}{2} + h)}$$

$$= C_{p,2} \frac{\Gamma(\frac{n.-1}{2} - \frac{p}{2} + h)}{\Gamma(\frac{n.-1}{2} + h)}$$

since the numerator gamma functions, except the last one, cancel with the denominator gamma functions except the first one. This expression happens to be the h-th moment of a real scalar type-1 beta random variable with the parameters $(\frac{n.-1-p}{2}, \frac{p}{2})$ and hence, for $k = 2$, $n. - 1 - p > 0$ and $p \geq 1$, w is a real scalar type-1 beta random variable. Then, we reject the null hypothesis H_o for small values of w or when the observed value of $w \leq w_\alpha$, with w_α such that $\int_0^{w_\alpha} f_w(w)dw = \alpha$ for a preassigned significance level α, $f_w(w)$ being the density of w for this case. We will use the same notation $f_w(w)$ for the density of w in all the special cases.

Case (4): $k = 3$, $p \geq 1$

Proceeding as in Case (3), we see that all the gammas in the h-th moment of w cancel out except the last two in the numerator and the first two in the denominator. Thus,

$$E[w^h] = C_{p,3} \frac{\Gamma(\frac{n.-3}{2} + \frac{1}{2} - \frac{p-1}{2} + h)\Gamma(\frac{n.-3}{2} - \frac{p-1}{2})}{\Gamma(\frac{n.-1}{2} + h)\Gamma(\frac{n.-1}{2} - \frac{1}{2} + h)}$$

$$= C_{p,3} \frac{\Gamma(\frac{n.-1}{2} - \frac{p}{2} + h)\Gamma(\frac{n.-1}{2} - \frac{p}{2} - \frac{1}{2} + h)}{\Gamma(\frac{n.-1}{2} + h)\Gamma(\frac{n.-1}{2} - \frac{1}{2} + h)}.$$

After combining the gammas in $y = w^{\frac{1}{2}}$ with the help of the duplication formula (13.3.12), we have the following:

$$E[y^h] = \frac{\Gamma(n_. - 2)}{\Gamma(n_. - 2 - p)} \frac{\Gamma(n_. - p - 2 + h)}{\Gamma(n_. - 2 + h)}.$$

Therefore, $y = w^{\frac{1}{2}}$ is a real scalar type-1 random variable with the parameters $(n_. - p - 2, p)$. We reject the null hypothesis for small values of y or when the observed value of $y \leq y_\alpha$, with y_α such that $\int_0^{y_\alpha} f_y(y) dy = \alpha$ for a preassigned significance level α. We will use the same notation $f_y(y)$ for the density of y in all special cases.

We can also obtain some special cases for $t_1 = \frac{1-w}{w}$ and $t_2 = \frac{1-y}{y}$, with $y = \sqrt{w}$. With this transformation, t_1 and t_2 will be available in terms of type-2 beta variables in the real scalar case, which conveniently enables us to relate this distribution to real scalar F random variables so that an F table can be used for testing the null hypothesis and reaching a decision. We have noted that

$$w = \frac{|V|}{|U + V|} = |(U + V)^{-\frac{1}{2}} V (U + V)^{-\frac{1}{2}}| = |W_1|$$

$$= \frac{1}{|V^{-\frac{1}{2}} U V^{-\frac{1}{2}} + I|} = \frac{1}{|W_2 + I|}$$

where W_1 is a real matrix-variate type-1 beta random variable with the parameters $(\frac{n_. - k}{2}, \frac{k-1}{2})$ and W_2 is a real matrix-variate type-2 beta random variable with the parameters $(\frac{k-1}{2}, \frac{n_. - k}{2})$. Then, when $p = 1$, W_1 and W_2 are real scalar variables, denoted by w_1 and w_2, respectively. Then for $p = 1$, we have one gamma ratio with h in the general h-th moment (13.3.7) and then,

$$t_1 = \frac{1 - w}{w} = \frac{1}{w} - 1 = (w_2 + 1) - 1 = w_2$$

where w_2 is a real scalar type-2 beta random variable with the parameters $(\frac{p-1}{2}, \frac{n_. - k}{2})$. As well, in general, for a real matrix-variate type-2 beta matrix W_2 with the parameters $(\frac{v_1}{2}, \frac{v_2}{2})$, we have $\frac{v_2}{v_1} W_2 = F_{v_1, v_2}$ where F_{v_1, v_2} is a real matrix-variate F matrix random variable with degrees of freedom v_1 and v_2. When $p = 1$ or in the real scalar case $\frac{v_2}{v_1} w_2 = F_{v_1, v_2}$ where, in this case, F is a real scalar F random variable with v_1 and v_2 degrees of freedom. We have used F for the scalar and matrix-variate case in order to avoid too many symbols. For $p = 2$, we combine the gamma functions in the numerator and denominator by applying the duplication formula for gamma functions (13.3.12); then, for $t_2 = \frac{1-y}{y}$ the situation turns out to be the same as in the case of t_1, the only difference being that in the

real scalar type-2 beta w_2, the parameters are $(k - 1, n_{\cdot} - k - 1)$. Note that the original $\frac{k-1}{2}$ has become $k - 1$ and the original $\frac{n_{\cdot}-k}{2}$ has become $n_{\cdot} - k - 1$. Thus, we can state the following two special cases.

Case (5): $p = 1$, $t_1 = \frac{1-w}{w}$

As was explained, t_1 is a real type-2 beta random variable with the parameters $(\frac{k-1}{2}, \frac{n_{\cdot}-k}{2})$, so that

$$\frac{n_{\cdot} - k}{k - 1} t_1 \simeq F_{k-1,n_{\cdot}-k},$$

which is a real scalar F random variable with $k - 1$ and $n_{\cdot} - k$ degrees of freedom. Accordingly, we reject H_o for small values of w and y, which corresponds to large values of F. Thus, we reject the null hypothesis H_o whenever the observed value of $F_{k-1,n_{\cdot}-k} \geq F_{k-1,n_{\cdot}-k,\alpha}$ where $F_{k-1,n_{\cdot}-k,\alpha}$ is the upper $100\,\alpha\%$ percentage point of the F distribution or $\int_a^\infty g(F)\mathrm{d}F = \alpha$ where $a = F_{k-1,n_{\cdot}-k,\alpha}$ and $g(F)$ is the density of F in this case.

Case (6): $p = 2$, $t_2 = \frac{1-y}{y}$, $y = \sqrt{w}$

As previously explained, t_2 is a real scalar type-2 beta random variable with the parameters $(k - 1, n_{\cdot} - k - 1)$ or

$$\frac{n_{\cdot} - k - 1}{k - 1} t_2 \simeq F_{2(k-1),2(n_{\cdot}-k-1)},$$

which is a real scalar F random variable having $2(k - 1)$ and $2(n_{\cdot} - k - 1)$ degrees of freedom. We reject the null hypothesis for large values of t_2 or when the observed value of $[\frac{n_{\cdot}-k-1}{k-1}]t_2 \geq b$ with b such that $\int_b^\infty g(F)\mathrm{d}F = \alpha$, $g(F)$ denoting in this case the density of a real scalar random variable F with degrees of freedoms $2(k - 1)$ and $2(n_{\cdot} - k - 1)$, and $b = F_{2(k-1),2(n_{\cdot}-k-1),\alpha}$.

Case (7): $k = 2$, $p \geq 1$, $t_1 = \frac{1-w}{w}$

For the case $k = 2$, we have already seen that the gamma functions with h in their arguments cancel out, leaving only one gamma in the numerator and one gamma in the denominator, so that w is distributed as a real scalar type-1 beta random variable with the parameters $(\frac{n_{\cdot}-1-p}{2}, \frac{p}{2})$. Thus, $t_1 = \frac{1-w}{w}$ is a real scalar type-2 beta with the parameters $(\frac{p}{2}, \frac{n_{\cdot}-p-1}{2})$, and

$$\left[\frac{n_{\cdot} - 1 - p}{p}\right]t_1 \simeq F_{p,n_{\cdot}-1-p},$$

which is a real scalar F random variable having p and $n_{\cdot} - 1 - p$ degrees of freedom. We reject H_o for large values of t_1 or when the observed value of $[\frac{n_{\cdot}-1-p}{p}]t_1 \geq b$ where b

is such that $\int_b^\infty g(F)\mathrm{d}F = \alpha$ with $g(F)$ being the density of an F random variable with degrees of freedoms p and $n_. - 1 - p$ in this special case.

Case (8): $k = 3$, $p \geq 1$, $t_2 = \frac{1-y}{y}$, $y = \sqrt{w}$

On combining Cases (4) and (6), it is seen that t_2 is a real scalar type-2 beta random variable with the parameters $(p, n_. - p - 1)$, so that

$$\frac{n_. - 1 - p}{p} t_2 \simeq F_{2p,2(n_.-p-1)},$$

which is a real scalar F random variable with the degrees of freedoms $(2p, 2(n_. - p - 1))$. Thus, we reject the hypothesis for large values of this F random variable. For a test at significance level α or with α as the size of its critical region, the hypothesis $H_o : A_1 = A_2 = \cdots = A_k = O$ is rejected when the observed value of this $F \geq F_{2p,2(n_.-1-p),\alpha}$ where $F_{2p,2(n_.-1-p),\alpha}$ is the upper $100\,\alpha\%$ percentage point of the F distribution.

Example 13.3.1. In a dieting experiment, three different diets D_1, D_2 and D_3 are tried for a period of one month. The variables monitored are weight in kilograms (kg), waist circumference in centimeters (cm) and right mid-thigh circumference in centimeters. The measurements are $x_1 = $ final weight minus initial weight, $x_2 = $ final waist circumference minus initial waist reading and $x_3 = $ final minus initial thigh circumference. Diet D_1 is administered to a group of 5 randomly selected individuals ($n_1 = 5$), D_2, to 4 randomly selected persons ($n_2 = 4$), and 6 randomly selected individuals ($n_3 = 6$) are subjected to D_3. Since three variables are monitored, $p = 3$. As well, there are three treatments or three diets, so that $k = 3$. In our notation,

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad X_{ij} = \begin{bmatrix} x_{1ij} \\ x_{2ij} \\ x_{3ij} \end{bmatrix}, \quad X_{1j} = \begin{bmatrix} x_{11j} \\ x_{21j} \\ x_{31j} \end{bmatrix}, \quad X_{2j} = \begin{bmatrix} x_{12j} \\ x_{22j} \\ x_{32j} \end{bmatrix}, \quad X_{3j} = \begin{bmatrix} x_{13j} \\ x_{23j} \\ x_{33j} \end{bmatrix},$$

where i corresponds to the diet number and j stands for the sample serial number. For example, the observation vector on individual #3 within the group subjected to diet D_2 is denoted by X_{23}. The following are the data on x_1, x_2, x_3:

Diet $D_1 : X_{1j}$, $j = 1, 2, 3, 4, 5$:

$$X_{11} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \quad X_{12} = \begin{bmatrix} 4 \\ -2 \\ -1 \end{bmatrix}, \quad X_{13} = \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}, \quad X_{14} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}, \quad X_{15} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Diet D_2 : $X_{2j}, \; j = 1, 2, 3, 4$:

$$X_{21} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \quad X_{22} = \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix}, \quad X_{23} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad X_{24} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

Diet D_3 : $X_{3j}, \; j = 1, 2, 3, 4, 5, 6$:

$$X_{31} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \quad X_{32} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \quad X_{33} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix},$$

$$X_{34} = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}, \quad X_{35} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \quad X_{36} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

(1): Perform an ANOVA test on the first component consisting of weight measurements; (2): Carry out a MANOVA test on the first two components, weight and waist measurements; (3): Do a MANOVA test on all the three variables, weight, waist and thigh measurements.

Solution 13.3.1.   We first compute the vectors $X_{1.}, \bar{X}_1, X_{2.}, \bar{X}_2, X_{3.}, \bar{X}_3, X_{..}$ and $\bar{X}$:

$$X_{1.} = \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}, \; \bar{X}_1 = \frac{X_{1.}}{n_1} = \frac{1}{5}\begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; X_{2.} = \begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix}, \; \bar{X}_2 = \frac{X_{2.}}{n_2} = \frac{1}{4}\begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},$$

$$X_{3.} = \begin{bmatrix} 6 \\ 12 \\ 6 \end{bmatrix}, \; \bar{X}_3 = \frac{X_{3.}}{6} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \; X_{..} = \begin{bmatrix} 15 \\ 16 \\ 6 \end{bmatrix}, \; \bar{X} = \frac{X_{..}}{n_.} = \frac{1}{15}\begin{bmatrix} 15 \\ 16 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 16/15 \\ 6/15 \end{bmatrix}.$$

Problem (1): ANOVA on the first component x_1.   The first components of the observations are x_{1ij}. The first components under diet D_1 are

$$[x_{111}, x_{112}, x_{113}, x_{114}, x_{115}] = [2, 4, -1, -1, 1] \text{ with } x_{11.} = 5;$$

the first components of observations under diet D_2 are

$$[x_{121}, x_{122}, x_{123}, x_{124}] = [1, 3, -1, 1] \text{ with } x_{12.} = 4;$$

and the first components under diet D_3 are

$$[x_{131}, x_{132}, x_{133}, x_{134}, x_{135}, x_{136}] = [2, 1, -1, 2, 2, 0] \text{ with } x_{13.} = 6.$$

Hence, the total on the first component $x_{1..} = 15$, and $\bar{x}_1 = \frac{x_{1..}}{n_.} = \frac{15}{15} = 1$. The first component model is the following:

$$x_{1ij} = \mu + \alpha_i + e_{1ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, k.$$

Note again that estimators and estimates will be denoted by a hat. As previously mentioned, the same symbols will be used for the variables and the observations on those variables in order to avoid using too many symbols; however, the notations will be clear from the context. If the discussion pertains to distributions, then variables are involved, and if we are referring to numbers, then we are dealing with observations.

The least squares estimates are $\hat{\mu} = \frac{x_{1..}}{n_.} = 1$, $\hat{\alpha}_1 = \frac{x_{11.}}{5} = \frac{5}{5} = 1$, $\hat{\alpha}_2 = \frac{x_{12.}}{4} = \frac{4}{4} = 1$, $\hat{\alpha}_3 = \frac{x_{13.}}{6} = \frac{6}{6} = 1$. The first component hypothesis is $\alpha_1 = \alpha_2 = \alpha_3 = 0$. The total sum of squares is

$$\sum_{ij}(x_{1ij} - \bar{x}_1)^2 = \sum_{ij} x_{1ij}^2 - \frac{x_{1..}^2}{n_.}$$

$$= (2-1)^2 + (4-1)^2 + (-1-1)^2 + (-1-1)^2 + (1-1)^2$$
$$+ (1-1)^2 + (3-1)^2 + (-1-1)^2 + (1-1)^2$$
$$+ (2-1)^2 + (1-1)^2 + (-1-1)^2 + (2-1)^2 + (2-1)^2 + (0-1)^2$$
$$= 34.$$

The sum of squares due to the α_i's is available from

$$\sum_i n_i \left(\frac{x_{1i.}}{n_i} - \frac{x_{1..}}{n_.}\right)^2 = \sum_i \frac{x_{1i.}^2}{n_i} - \frac{x_{1..}^2}{n_.}$$

$$= 5\left(\frac{5}{5} - \frac{15}{15}\right)^2 + 4\left(\frac{4}{4} - \frac{15}{15}\right)^2 + 6\left(\frac{6}{6} - \frac{15}{15}\right)^2 = 0.$$

Hence the following table:

ANOVA Table

Variation due to	df	SS	MS	F-ratio
diets	2	0	0	0
residuals	12	34		
total	14	34		

Since the sum of squares due to the α_i's is null, the hypothesis is not rejected at any level.

Problem (2): MANOVA on the first two components. We are still using the notation X_{ij} for the two and three-component cases since our general notation does not depend on the number of components in the vector concerned; as well, we can make use of the computations pertaining to the first component in Problem (1). The relevant quantities computed from the data on the first two components are the following:

$$\text{Diet } D_1: \begin{bmatrix} 2 & 4 & -1 & -1 & 1 \\ 3 & -2 & -2 & 1 & 0 \end{bmatrix}, \ X_{1.} = \begin{bmatrix} x_{11.} \\ x_{21.} \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}, \ \bar{X}_1 = \frac{1}{5}\begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix};$$

$$\text{Diet } D_2: \begin{bmatrix} 1 & 3 & -1 & 1 \\ 2 & -1 & 2 & 1 \end{bmatrix}, \ X_{2.} = \begin{bmatrix} x_{12.} \\ x_{22.} \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}, \ \bar{X}_2 = \frac{1}{4}\begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix};$$

$$\text{Diet } D_3: \begin{bmatrix} 2 & 1 & -1 & 2 & 2 & 0 \\ 2 & 3 & 2 & 4 & 0 & 1 \end{bmatrix}, \ X_{3.} = \begin{bmatrix} 6 \\ 12 \end{bmatrix}, \ \bar{X}_3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

In this case, the grand total, denoted by $X_{..}$, and the grand average, denoted by $\bar{X}$, are the following:

$$X_{..} = \begin{bmatrix} 15 \\ 16 \end{bmatrix}, \ \bar{X} = \frac{1}{15}\begin{bmatrix} 15 \\ 16 \end{bmatrix} = \begin{bmatrix} 1 \\ 16/15 \end{bmatrix}.$$

Note that the total sum of squares and cross products matrix can be written as follows:

$$\sum_{ij}(X_{ij} - \bar{X})(X_{ij} - \bar{X})' = \sum_{j=1}^{5}(X_{1j} - \bar{X})(X_{1j} - \bar{X})' + \sum_{j=1}^{4}(X_{2j} - \bar{X})(X_{2j} - \bar{X})'$$

$$+ \sum_{j=1}^{6}(X_{3j} - \bar{X})(X_{3j} - \bar{X})'.$$

Then,

$$\sum_{j=1}^{5}(X_{1j} - \bar{X})(X_{1j} - \bar{X})' = \begin{bmatrix} 1 & 29/15 \\ 29/15 & 29^2/15^2 \end{bmatrix} + \begin{bmatrix} 9 & -138/15 \\ -138/15 & 46^2/15^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 4 & 92/15 \\ 92/15 & 46^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & 2/15 \\ 2/15 & 1/15^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 16^2/15^2 \end{bmatrix}$$

$$= \begin{bmatrix} 18 & -1 \\ -1 & 5330/15^2 \end{bmatrix};$$

$$\sum_{j=1}^{4}(X_{2j} - \bar{X})(X_{2j} - \bar{X})' = \begin{bmatrix} 0 & 0 \\ 0 & 14^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & -62/15 \\ -62/15 & 31^2/15^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 4 & -28/15 \\ -28/15 & 14^2/15^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1/15^2 \end{bmatrix} = \begin{bmatrix} 8 & -6 \\ -6 & 1354/15^2 \end{bmatrix};$$

$$\sum_{j=1}^{6}(X_{3j} - \bar{X})(X_{3j} - \bar{X})' = \begin{bmatrix} 1 & 14/15 \\ 14/15 & 14^2/15^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 29^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & -28/15 \\ -28/15 & 14^2/15^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 1 & 44/15 \\ 44/15 & 44^2/15^2 \end{bmatrix} + \begin{bmatrix} 1 & -16/15 \\ -16/15 & 16^2/15^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 1 & 1/15 \\ 1/15 & 1/15^2 \end{bmatrix} = \begin{bmatrix} 8 & 1 \\ 1 & 3426/15^2 \end{bmatrix};$$

$$\sum_{ij}(X_{ij} - \bar{X})(X_{ij} - \bar{X})' = \begin{bmatrix} 18 & -1 \\ -1 & 5330/15^2 \end{bmatrix} + \begin{bmatrix} 8 & -6 \\ -6 & 1354/15^2 \end{bmatrix} + \begin{bmatrix} 8 & 1 \\ 1 & 3426/15^2 \end{bmatrix}$$

$$= \begin{bmatrix} 34 & -6 \\ -6 & 674/15 \end{bmatrix} \text{ and } \begin{vmatrix} 34 & -6 \\ -6 & 674/15 \end{vmatrix} = 1491.73.$$

Now, consider the residual sum of squares and cross products matrix:

$$\sum_{ij}(X_{ij} - \tfrac{X_{i.}}{n_i})(X_{ij} - \tfrac{X_{i.}}{n_i})' = \sum_{j=1}^{5}(X_{1j} - \tfrac{X_{1.}}{n_1})(X_{1j} - \tfrac{X_{1.}}{n_1})'$$

$$+ \sum_{j=1}^{4}(X_{2j} - \tfrac{X_{2.}}{n_2})(X_{2j} - \tfrac{X_{2.}}{n_2})' + \sum_{j=1}^{6}(X_{3j} - \tfrac{X_{3.}}{n_3})(X_{3j} - \tfrac{X_{3.}}{n_3})'.$$

That is,

$$\sum_{j=1}^{5}(X_{1j} - \tfrac{X_{1.}}{n_1})(X_{1j} - \tfrac{X_{1.}}{n_1})' = \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix} + \begin{bmatrix} 9 & -6 \\ -6 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

$$+ \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 18 & -1 \\ -1 & 18 \end{bmatrix};$$

$$\sum_{j=1}^{4}(X_{2j} - \tfrac{X_{2.}}{n_2})(X_{2j} - \tfrac{X_{2.}}{n_2})' = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} + \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 8 & -6 \\ -6 & 6 \end{bmatrix};$$

$$\sum_{j=1}^{6}(X_{3j} - \tfrac{X_{3.}}{n_3})(X_{3j} - \tfrac{X_{3.}}{n_3})' = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

$$+ \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 1 \\ 1 & 10 \end{bmatrix}.$$

Hence,

$$\sum_{ij}(X_{ij} - \tfrac{X_{i.}}{n_i})(X_{ij} - \tfrac{X_{i.}}{n_i})' = \begin{bmatrix} 18 & -1 \\ -1 & 18 \end{bmatrix} + \begin{bmatrix} 8 & -6 \\ -6 & 6 \end{bmatrix} + \begin{bmatrix} 8 & 1 \\ 1 & 10 \end{bmatrix}$$

$$= \begin{bmatrix} 34 & -6 \\ -6 & 34 \end{bmatrix} \text{ and } \begin{vmatrix} 34 & -6 \\ -6 & 34 \end{vmatrix} = 1120.$$

Therefore, the observed w is given by

$$w = \frac{1120}{1491.73} = 0.7508, \ \sqrt{w} = 0.8665.$$

This is the case $p = 2$, $k = 3$, that is, our special Case (8). Then, the observed value of

$$t_2 = \frac{1 - \sqrt{w}}{\sqrt{w}} = \frac{0.1335}{0.8665} = 0.1540,$$

and

$$\frac{n_. - 1 - p}{p} t_2 = \frac{15 - 1 - 2}{2}(0.1540) = 0.9244.$$

Our F-statistic is $F_{2p, 2(n_. - p - 1)} = F_{4,24}$. Let us test the hypothesis $A_1 = A_2 = A_3 = O$ at the 5% significance level or $\alpha = 0.05$. Since the observed value $0.9244 < 5.77 = F_{4,24,0.05}$ which is available from F-tables, we do not reject the hypothesis.

Verification of the calculations

Denoting the total sum of squares and cross products matrix by S_t, the residual sum of squares and cross products matrix by S_r and the sum of squares and cross products matrix due to the hypothesis or due to the effects A_i's by S_h, we should have $S_t = S_r + S_h$ where

$$S_t = \begin{bmatrix} 34 & -6 \\ -6 & 674/15 \end{bmatrix} \text{ and } S_r = \begin{bmatrix} 34 & -6 \\ -6 & 34 \end{bmatrix}$$

as previously determined. Let us compute

$$S_h = \sum_{i=1}^{k} n_i \left(\frac{X_{i.}}{n_i} - \frac{X_{..}}{n_.} \right) \left(\frac{X_{i.}}{n_i} - \frac{X_{..}}{n_.} \right)'.$$

For the first two components, we already have the following:

$$\frac{X_{1.}}{n_1} - \frac{X_{..}}{n_.} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 16/15 \end{bmatrix} = \begin{bmatrix} 0 \\ -16/15 \end{bmatrix}, \; n_1 = 5$$

$$\frac{X_{2.}}{n_2} - \frac{X_{..}}{n_.} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 16/15 \end{bmatrix} = \begin{bmatrix} 0 \\ -1/15 \end{bmatrix}, \; n_2 = 4$$

$$\frac{X_{3.}}{n_3} - \frac{X_{..}}{n_.} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 16/15 \end{bmatrix} = \begin{bmatrix} 0 \\ 14/15 \end{bmatrix}, \; n_3 = 6.$$

Hence,

$$S_h = 5 \begin{bmatrix} 0 & 0 \\ 0 & 16^2/15^2 \end{bmatrix} + 4 \begin{bmatrix} 0 & 0 \\ 0 & 1/15^2 \end{bmatrix} + 6 \begin{bmatrix} 0 & 0 \\ 0 & 14^2/15^2 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & 2460/15^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 164/15 \end{bmatrix}.$$

As $34 + \frac{164}{15} = \frac{674}{15}$, $S_t = S_r + S_h$, that is,

$$\begin{bmatrix} 34 & -6 \\ -6 & 674/15 \end{bmatrix} = \begin{bmatrix} 34 & -6 \\ -6 & 34 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 164/15 \end{bmatrix}.$$

Thus, the result is verified.

Problem (3): Data on all the three variables. In this case, we have $p = 3, k = 3$. We will first use $X_{1.}, \bar{X}_1, X_{2.}, \bar{X}_2, X_{3.}, \bar{X}_3, X_{..}$ and $\bar{X}$ which have already been evaluated, to compute the residual sum of squares and cross product matrix. Since all the matrices are symmetric, for convenience, we will only display the diagonal elements and those above the diagonal. As in the case of two components, we compute the following, making use of the calculations already done for the 2-component case (the notations remaining the same since our general notation does not involve p):

$$\sum_{j=1}^{5}(X_{1j} - \tfrac{X_{1.}}{n_1})(X_{1j} - \tfrac{X_{1.}}{n_1})' = \begin{bmatrix} 1 & 3 & 1 \\ & 9 & 3 \\ & & 1 \end{bmatrix} + \begin{bmatrix} 9 & -6 & -3 \\ & 4 & 2 \\ & & 1 \end{bmatrix} + \begin{bmatrix} 4 & 4 & -2 \\ & 4 & -2 \\ & & 1 \end{bmatrix}$$

$$+ \begin{bmatrix} 4 & -2 & 2 \\ & 1 & -1 \\ & & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ & 0 & 0 \\ & & 0 \end{bmatrix} = \begin{bmatrix} 18 & -1 & -2 \\ & 18 & 2 \\ & & 4 \end{bmatrix} ;$$

$$\sum_{j=1}^{4}(X_{2j} - \tfrac{X_{2.}}{n_2})(X_{2j} - \tfrac{X_{2.}}{n_2})' = \begin{bmatrix} 0 & 0 & 0 \\ & 1 & 2 \\ & & 4 \end{bmatrix} + \begin{bmatrix} 4 & -4 & -4 \\ & 4 & 4 \\ & & 4 \end{bmatrix} + \begin{bmatrix} 4 & -2 & -2 \\ & 1 & 1 \\ & & 1 \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 & 0 \\ & 0 & 0 \\ & & 1 \end{bmatrix} = \begin{bmatrix} 8 & -6 & -6 \\ & 6 & 7 \\ & & 10 \end{bmatrix} ;$$

$$\sum_{j=1}^{6}(X_{3j} - \tfrac{X_{3.}}{n_3})(X_{3j} - \tfrac{X_{3.}}{n_3})' = \begin{bmatrix} 1 & 0 & 2 \\ & 0 & 0 \\ & & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ & 1 & 0 \\ & & 0 \end{bmatrix} + \begin{bmatrix} 4 & 0 & -2 \\ & 0 & 0 \\ & & 1 \end{bmatrix}$$

$$+ \begin{bmatrix} 1 & 2 & 1 \\ & 4 & 2 \\ & & 1 \end{bmatrix} + \begin{bmatrix} 1 & -2 & -1 \\ & 4 & 2 \\ & & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 & -1 \\ & 1 & -1 \\ & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 8 & 1 & -5 \\ & 10 & 3 \\ & & 8 \end{bmatrix} .$$

Then,

$$\sum_{ij}(X_{ij} - \tfrac{X_{i.}}{n_i})(X_{ij} - \tfrac{X_{i.}}{n_i})' = \begin{bmatrix} 18 & -1 & -2 \\ & 18 & 2 \\ & & 4 \end{bmatrix} + \begin{bmatrix} 8 & -6 & -6 \\ & 6 & 7 \\ & & 10 \end{bmatrix} + \begin{bmatrix} 8 & 1 & -5 \\ & 10 & 3 \\ & & 8 \end{bmatrix}$$

$$= \begin{bmatrix} 34 & -6 & -13 \\ & 34 & 12 \\ & & 22 \end{bmatrix}$$

whose determinant is equal to

$$= 34 \begin{vmatrix} 34 & 12 \\ 12 & 22 \end{vmatrix} + 6 \begin{vmatrix} -6 & 12 \\ -13 & 22 \end{vmatrix} - 13 \begin{vmatrix} -6 & 34 \\ -13 & 12 \end{vmatrix}$$

$$= 15870.$$

The total sum of squares and cross products matrix is the following:

$$\sum_{ij}(X_{ij}-\bar{X})(X_{ij}-\bar{X})' = \sum_{j=1}^{5}(X_{1j}-\bar{X})(X_{1j}-\bar{X})' + \sum_{j=1}^{4}(X_{2j}-\bar{X})(X_{2j}-\bar{X})'$$
$$+ \sum_{j=1}^{6}(X_{3j}-\bar{X})(X_{3j}-\bar{X})',$$

with

$$\sum_{j=1}^{5}(X_{1j}-\bar{X})(X_{1j}-\bar{X})' = \begin{bmatrix} 1 & 29/15 & 9/15 \\ & 29^2/15^2 & (29 \times 9)/15^2 \\ & & 9^2/15^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 9 & -(3 \times 46)/15 & -(3 \times 21)/15 \\ & 46^2/15^2 & (46 \times 21)/15^2 \\ & & 21^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & 92/15 & -18/15 \\ & 46^2/15^2 & -(46 \times 9)/15^2 \\ & & 9^2/15^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 4 & 2/15 & 42/15 \\ & 1/15^2 & 21/15^2 \\ & & 21^2/15^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ & 16^2/15^2 & 96/15^2 \\ & & 36/15^2 \end{bmatrix}$$

$$= \begin{bmatrix} 18 & -1 & -2 \\ & 5330/15^2 & 930/15^2 \\ & & 1080/15^2 \end{bmatrix},$$

$$\sum_{j=1}^{4}(X_{2j}-\bar{X})(X_{2j}-\bar{X})' = \begin{bmatrix} 0 & 0 & 0 \\ & 14^2/15^2 & (14 \times 24)/15^2 \\ & & 24^2/15^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 4 & -62/15 & -72/15 \\ & 31^2/15^2 & (31 \times 36)/15^2 \\ & & 36^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & -28/15 & -18/15 \\ & 14^2/15^2 & (14 \times 9)/15^2 \\ & & 9^2/15^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 & 0 \\ & 1/15^2 & 21/15^2 \\ & & 21^2/15^2 \end{bmatrix} = \begin{bmatrix} 8 & -6 & -6 \\ & 1354/15^2 & 1599/15^2 \\ & & 2394/15^2 \end{bmatrix},$$

$$\sum_{j=1}^{6}(X_{3j}-\bar{X})(X_{3j}-\bar{X})'=\begin{bmatrix}1 & 14/15 & -21/15\\ & 14^2/15^2 & -(14\times21)/15^2\\ & & 21^1/15^2\end{bmatrix}$$

$$+\begin{bmatrix}0 & 0 & 0\\ & 29^2/15^2 & (29\times9)/15^2\\ & & 9^2/15^2\end{bmatrix}+\begin{bmatrix}4 & -28/15 & -48/15\\ & 14^2/15^2 & (14\times24)/15^2\\ & & 24^2/15^2\end{bmatrix}$$

$$+\begin{bmatrix}1 & 44/15 & 24/15\\ & 44^2/15^2 & (44\times24)/15^2\\ & & 24^2/15^2\end{bmatrix}+\begin{bmatrix}1 & -16/15 & -6/15\\ & 16^2/15^2 & 96/15^2\\ & & 6^2/15^2\end{bmatrix}$$

$$+\begin{bmatrix}1 & 1/15 & -24/15\\ & 1/15^2 & -24/15^2\\ & & 24^2/15^2\end{bmatrix}=\begin{bmatrix}8 & 1 & -5\\ & 3426/15^2 & 1431/15^2\\ & & 2286/15^2\end{bmatrix}.$$

Hence the total sum of squares and cross products matrix is

$$\sum_{ij}(X_{ij}-\bar{X})(X_{ij}-\bar{X})'=\begin{bmatrix}18 & -1 & -2\\ & 5330/15^2 & 930/15^2\\ & & 1080/15^2\end{bmatrix}+\begin{bmatrix}8 & -6 & -6\\ & 1354/15^2 & 1599/15^2\\ & & 2394/15^2\end{bmatrix}$$

$$+\begin{bmatrix}8 & 1 & -5\\ & 3426/15^2 & 1431/15^2\\ & & 2286/15^2\end{bmatrix}=\begin{bmatrix}34 & -6 & -13\\ -6 & 674/15 & 264/15\\ -13 & 264/15 & 384/15\end{bmatrix}$$

and its determinant $=34\begin{vmatrix}674/15 & 264/15\\264/15 & 384/15\end{vmatrix}+6\begin{vmatrix}-6 & 264/15\\-13 & 384/15\end{vmatrix}$

$$-13\begin{vmatrix}-6 & 674/15\\-13 & 264/15\end{vmatrix}=342126/15.$$

Then, the observed value of

$$w=\frac{15870\times15}{342126}=0.6958,\quad\sqrt{w}=0.8341.$$

Since $p=3$ and $k=3$, an exact distribution is available from our special Case (8) for $t_2=\frac{1-\sqrt{w}}{\sqrt{w}}$ and an observed value of $t_2=0.1989$. Then,

$$\frac{n_.-1-p}{p}t_2=\frac{15-1-3}{3}t_2=\frac{11}{3}t_2\sim F_{2p,2(n_.-1-p)}=F_{6,22}.$$

The critical value obtained from an F-table at the 5% significance level is $F_{6,22,.05} \approx 3.85$. Since the observed value of $F_{6,22}$ is $\frac{11}{3}(0.1989) = 0.7293 < 3.85$, the hypothesis $A_1 = A_2 = A_3 = O$ is not rejected. It can also be verified that $S_t = S_r + S_h$.

13.3.4. Asymptotic distribution of the λ-criterion

We can obtain an asymptotic real chisquare distribution for $n_. \to \infty$. To this end, consider the general h-th moments of λ or $E[\lambda^h]$ from (13.3.8), that is,

$$E[\lambda^h] = C_{p,k} \prod_{j=1}^{p} \left[\Gamma\left(\frac{n_.-k}{2} - \frac{j-1}{2} + \frac{n_.}{2}h\right) \Big/ \Gamma\left(\frac{n_.-1}{2} - \frac{j-1}{2} + \frac{n_.}{2}h\right) \right]$$

$$= C_{p,k} \prod_{j=1}^{p} \left[\Gamma\left(\frac{n_.}{2}(1+h) - \frac{j-1}{2} - \frac{k}{2}\right) \Big/ \Gamma\left(\frac{n_.}{2}(1+h) - \frac{j-1}{2} - \frac{1}{2}\right) \right].$$

Let us expand all the gamma functions in $E[\lambda^h]$ by using the first term in the asymptotic expansion of a gamma function or by making use of Stirling's approximation formula, namely,

$$\Gamma(z + \delta) \approx \sqrt{(2\pi)} z^{z+\delta-\frac{1}{2}} e^{-z} \tag{13.3.13}$$

for $|z| \to \infty$ when δ is a bounded quantity. Taking $\frac{n_.}{2} \to \infty$ in the constant part and $\frac{n_.}{2}(1+h) \to \infty$ in the part containing h, we have

$$\Gamma\left(\frac{n_.}{2}(1+h) - \frac{j-1}{2} - \frac{k}{2}\right) \Big/ \Gamma\left(\frac{n_.}{2}(1+h) - \frac{j-1}{2} - \frac{1}{2}\right)$$

$$\approx \left\{ \sqrt{(2\pi)}\left[\frac{n_.}{2}(1+h)\right]^{\frac{n_.}{2}(1+h) - \frac{j-1}{2} - \frac{k}{2} - \frac{1}{2}} e^{-\frac{n_.}{2}(1+h)} \right.$$

$$\left. \Big/ \sqrt{(2\pi)}\left[\frac{n_.}{2}(1+h)\right]^{\frac{n_.}{2}(1+h) - \frac{j-1}{2} - \frac{1}{2} - \frac{1}{2}} e^{\frac{n_.}{2}(1+h)} \right\}$$

$$= \left(\frac{n_.}{2}\right)^{-(\frac{k-1}{2})}(1+h)^{-(\frac{k-1}{2})}.$$

The factor $\left(\frac{n_.}{2}\right)^{-(\frac{k-1}{2})}$ is canceled from the expression coming from the constant part. Then, taking the product over $j = 1, \ldots, p$, we have

$$\lambda^h \to (1+h)^{-p(k-1)/2} \text{ or } \lambda^{-2h} \to (1-2h)^{-p(k-1)/2} \text{ for } 1 - 2h > 0,$$

which is the moment generating function (mgf) of a real scalar chisquare with $p(k-1)$ degrees of freedom. Hence, we have the following result:

Theorem 13.3.1. *Letting λ be the likelihood ratio criterion for testing the hypothesis $H_o : A_1 = A_2 = \cdots = A_k = O$, the asymptotic distribution of $-2 \ln \lambda$ is a real chisquare random variable having $p(k-1)$ degrees of freedom as $n_{\cdot} \to \infty$, that is,*

$$-2 \ln \lambda \to \chi^2_{p(k-1)} \ as \ n_{\cdot} \to \infty. \tag{13.3.14}$$

Observe that we only require the sum of the sample sizes $n_1 + \cdots + n_k = n_{\cdot}$ to go to infinity, and not that the individual n_j's be large. This chisquare approximation can be utilized for testing the hypothesis for large values of $n_{\cdot}$, and we then reject H_o for small values of λ, which means for large values of $-2 \ln \lambda$ or large values of $\chi^2_{p(k-1)}$, that is, when the observed $-2 \ln \lambda \geq \chi^2_{p(k-1),\alpha}$ where $\chi^2_{p(k-1),\alpha}$ denotes the upper $100\,\alpha\%$ percentage point of the chisquare distribution.

13.3.5. MANOVA and testing the equality of population mean values

In a one-way classification model, we have the following for the p-variate case:

$$X_{ij} = M + A_i + E_{ij} \ or \ X_{ij} = M_i + E_{ij}, \ with \ M_i = M + A_i, \tag{13.3.15}$$

for $j = 1, \ldots, n_i$, $i = 1, \ldots, k$. When the error vector is assumed to have a null expected value, that is, $E[E_{ij}] = O$, for all i and j, we have $E[X_{ij}] = M_i$ for all i and j. Thus, this assumption, in conjunction with the hypothesis $A_1 = A_2 = \cdots = A_k = O$, implies that $M_1 = M_2 = \cdots = M_k$, that is, the hypothesis of equality of the population mean value vectors or the test is equivalent to testing the equality of population mean value vectors in k independent populations with common covariance matrix $\Sigma > O$. We have already tackled this problem in Chap. 6 under both assumptions that Σ is known and unknown, when the populations are Gaussian, that is, $X_{ij} \sim N_p(M_i, \Sigma)$, $\Sigma > O$. Thus, the hypothesis made in a one-way classification MANOVA setting and the hypothesis of testing the equality of mean value vectors in MANOVA are one and the same. In the scalar case too, the ANOVA in a one-way classification data coincides with testing the equality of population mean values in k independent univariate populations. In the ANOVA case, we are comparing the sum of squares attributable to the hypothesis to the residual sum of squares. If the hypothesis really holds true, then the sum of squares due to the hypothesis or to the α_j's (deviations from the general effect due to the j-th treatment) must be zero and hence for large values of the sum of squares due to the presence of the α_j's, as compared to the residual sum of squares, we reject the hypothesis. In MANOVA, we are comparing two sums of squares and cross product matrices, namely,

$$U = \sum_{ij} \left[\frac{X_{i.}}{n_i} - \frac{X_{..}}{n_{\cdot}} \right]\left[\frac{X_{i.}}{n_i} - \frac{X_{..}}{n_{\cdot}} \right]' \ and \ V = \sum_{ij} \left[X_{ij} - \frac{X_{i.}}{n_i} \right]\left[X_{ij} - \frac{X_{i.}}{n_i} \right]'.$$

We have the following distributional properties:

$$T_1 = (U + V)^{-\frac{1}{2}} U (U + V)^{-\frac{1}{2}} \sim \text{ real matrix-variate type-1 beta with parameters}$$
$$(\tfrac{k-1}{2}, \tfrac{n.-k}{2});$$
$$T_2 = (U + V)^{-\frac{1}{2}} V (U + V)^{-\frac{1}{2}} \sim \text{ real matrix-variate type-1 beta with parameters}$$
$$(\tfrac{n.-k}{2}, \tfrac{k-1}{2});$$
$$T_3 = V^{-\frac{1}{2}} U V^{-\frac{1}{2}} \sim \text{ real matrix-variate type-2 beta with parameters } (\tfrac{k-1}{2}, \tfrac{n.-k}{2});$$
$$T_4 = U^{-\frac{1}{2}} V U^{-\frac{1}{2}} \sim \text{ real matrix-variate type-2 beta with parameters } (\tfrac{n.-k}{2}, \tfrac{k-1}{2}).$$
$$(13.3.16)$$

The likelihood ratio criterion is

$$\lambda = \frac{|V|}{|U + V|} = |T_2| = \frac{1}{|T_3 + I|} = \frac{1}{\prod_{j=1}^{p}(1 + \eta_j)} \tag{13.3.17}$$

where the η_j's are the eigenvalues of T_3. We reject H_o for small values of λ which means for large values of $\prod_{j=1}^{p}[1 + \eta_j]$. The basic objective in MANOVA consists of comparing U and V, the matrices due to the presence of treatment effects and due to the residuals, respectively. We can carry out this comparison by using the type-1 beta matrices T_1 and T_2 or the type-2 beta matrices T_3 and T_4 or by making use of the eigenvalues of these matrices. In the type-1 beta case, the eigenvalues will be between 0 and 1, whereas in the type-2 beta case, the eigenvalues will be real positive or simply positive. We may also note that the eigenvalues of T_1 and its nonsymmetric forms $U(U + V)^{-1}$ or $(U + V)^{-1}U$ are identical. Similarly, the eigenvalues of the symmetric form T_2 and $V(U+V)^{-1}$ or $(U + V)^{-1}V$ are one and the same. As well, the eigenvalues of the symmetric form T_3 and the nonsymmetric forms UV^{-1} or $V^{-1}U$ are the same. Again, the eigenvalues of the symmetric form T_4 and its nonsymmetric forms $U^{-1}V$ or VU^{-1} are the same. Several researchers have constructed tests based on the matrices T_1, T_2, T_3, T_4 or their nonsymmetric forms or their eigenvalues. Some of the well-known test statistics are the following:

$$\text{Lawley-Hotelling trace} = \text{tr}(T_3)$$
$$\text{Roy's largest root} = \text{the largest eigenvalue of } T_2$$
$$\text{Pillai's trace} = \text{tr}(T_1)$$
$$\text{Wilks' lambda} = |T_2| = \text{ the likelihood ratio statistic.}$$

For example, when the hypothesis is true, we expect the eigenvalues of T_3 to be small and hence we may reject the hypothesis when its smallest eigenvalue is large or the trace of T_3 is large. If we are using T_4, then when the hypothesis is true, we expect T_4 to be large in the sense that the eigenvalues will be large, and therefore we may reject the hypothesis for small values of its largest eigenvalue or its trace. If we are utilizing T_1, we are actually comparing the contribution attributable to the treatments to the total variation. We expect this to be small under the hypothesis and hence, we may reject the hypothesis for large values of its smallest eigenvalue or its trace. If we are using T_2, we are comparing the residual part to the total variation. If the hypothesis is true, then we can expect a substantial contribution from the residual part so that we may reject the hypothesis for small values of the largest eigenvalue or the trace in this case. These are the main ideas in connection with constructing statistics for testing the hypothesis on the basis of the eigenvalues of the matrices T_1, T_2, T_3 and T_4.

13.3.6. When H_o is rejected

When $H_o : A_1 = \cdots = A_k = O$ is rejected, it is plausible that some of the differences may be non-null, that is, $A_i - A_j \neq O$ for some i and j, $i \neq j$. We may then test individual hypotheses of the type $H_{o1} : A_i = A_j$ for $i \neq j$. There are $k(k-1)/2$ such differences. This type of test is equivalent to testing the equality of the mean value vectors in two independent p-variate Gaussian populations with the same covariance matrix $\Sigma > O$. This has already been discussed in Chap. 6 for the cases Σ known and Σ unknown. In this instance, we can use the special Case (7) where for $k = 2$, and the statistic t_1 is real scalar type-2 beta distributed with the parameters $(\frac{p}{2}, \frac{n.-1-p}{2})$, so that

$$\frac{n. - 1 - p}{p} t_1 \sim F_{p, n. -1-p} \tag{13.3.18}$$

where $n. = n_i + n_j$ for some specific i and j. We can make use of (13.3.18) for testing individual hypotheses. By utilizing Special Case (8) for $k = 3$, we can also test a hypothesis of the type $A_i = A_j = A_m$ for different i, j, m. Instead of comparing the results of all the $k(k-1)/2$ individual hypotheses, we may examine the estimates of A_i, namely, $\hat{A}_i = \frac{X_{i.}}{n_i} - \frac{X_{..}}{n.}$, $i = 1, \ldots, k$. Consider the norms $\|\frac{X_{i.}}{n_i} - \frac{X_{j.}}{n_j}\|$, $i \neq j$ (the Euclidean norm may be taken for convenience). Start with the individual test corresponding to the maximum value of these norms. If this test is not rejected, it is likely that tests on all other differences will not be rejected either. If it is rejected, we then take the next largest difference and continue testing.

Note 13.3.1. Usually, before initiating a MANOVA, the assumption that the covariance matrices associated with the k populations or treatments are equal is tested. It may happen that the error variable E_{1j}, $j = 1, \ldots, n_1$, may have the common covariance matrix Σ_1, E_{2j}, $j = 1, \ldots, n_2$, may have the common covariance matrix Σ_2, and so on, where not all the Σ_j's equal. In this instance, we may first test the hypothesis $H_o : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_k$. This test is already described in Chap. 6. If this hypothesis is not rejected, we may carry out the MANOVA analysis of the data. If this hypothesis is rejected, then some of the Σ_j's may not be equal. In this case, we test individual hypotheses of the type $\Sigma_i = \Sigma_j$ for some specific i and j, $i \neq j$. Include all treatments for which the individual hypotheses are not rejected by the tests and exclude the data on the treatments whose Σ_j's may be different, but distinct from those already selected. Continue with the MANOVA analysis of the data on the treatments which are retained, that is, those for which the Σ_j's are equal in the sense that the corresponding tests of equality of covariance matrices did not reject the hypotheses.

Example 13.3.2. For the sake of illustration, test the hypothesis $H_o : A_1 = A_2$ with the data provided in Example 13.3.1.

Solution 13.3.2. We can utilize some of the computations done in the solution to Example 13.3.1. Here, $n_1 = 5, n_2 = 4$ and $n_. = n_1 + n_2 = 9$. We disregard the third sample. The residual sum of squares and cross products matrix in the present case is available from the Solution 13.3.1 by omitting the matrix corresponding to the third sample. Then,

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(X_{ij} - \frac{X_{i.}}{n_i}\right)\left(X_{ij} - \frac{X_{i.}}{n_i}\right)' = \begin{bmatrix} 18 & -1 & -2 \\ & 18 & 2 \\ & & 4 \end{bmatrix} + \begin{bmatrix} 8 & -6 & -6 \\ & 6 & 7 \\ & & 10 \end{bmatrix}$$

$$= \begin{bmatrix} 26 & -7 & -8 \\ -7 & 24 & 9 \\ -8 & 9 & 14 \end{bmatrix}$$

whose determinant is

$$26\begin{vmatrix} 24 & 9 \\ 9 & 14 \end{vmatrix} + 7\begin{vmatrix} -7 & 9 \\ -8 & 14 \end{vmatrix} - 8\begin{vmatrix} -7 & 24 \\ -8 & 9 \end{vmatrix} = 5416.$$

Let us compute $\sum_{i=1}^{2} \sum_{j=1}^{n_i} (X_{ij} - \frac{X_{..}}{n_.})(X_{ij} - \frac{X_{..}}{n_.})'$:

$$\sum_{j=1}^{5} \left(X_{1j} - \frac{X_{..}}{n_.}\right)\left(X_{1j} - \frac{X_{..}}{n_.}\right)' = \begin{bmatrix} 1 & 23/9 & 5/9 \\ & 23^2/9^2 & 115/9^2 \\ & & 5^2/9^2 \end{bmatrix} + \begin{bmatrix} 9 & -66/9 & -39/9 \\ & 22^2/9^2 & 286/9^2 \\ & & 13^2/9^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 4 & 22/9 & -10/9 \\ & 22^2/9^2 & -110/9^2 \\ & & 5^2/9^2 \end{bmatrix} + \begin{bmatrix} 4 & -10/9 & 26/9 \\ & 5^2/9^2 & -65/9^2 \\ & & 13^2/9^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 & 0 \\ & 4^2/9^2 & 16/9^2 \\ & & 4^2/9^2 \end{bmatrix} = \begin{bmatrix} 18 & -31/9 & -18/9 \\ & 1538/9^2 & 242/9^2 \\ & & 404/9^2 \end{bmatrix};$$

$$\sum_{j=1}^{4} \left(X_{2j} - \frac{X_{..}}{n_.}\right)\left(X_{2j} - \frac{X_{..}}{n_.}\right)' = \begin{bmatrix} 0 & 0 & 0 \\ & 14^2/9^2 & 14^2/9^2 \\ & & 14^2/9^2 \end{bmatrix} + \begin{bmatrix} 4 & -26/9 & -44/9 \\ & 13^2/9^2 & (13 \times 22)/9^2 \\ & & 22^2/9^2 \end{bmatrix}$$

$$+ \begin{bmatrix} 4 & -28/9 & -10/9 \\ & 14^2/9^2 & 70/9^2 \\ & & 5^2/9^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ & 5^2/9^2 & -65/9^2 \\ & & 13^2/9^2 \end{bmatrix}$$

$$= \begin{bmatrix} 8 & -6 & -6 \\ & 586/9^2 & 487/9^2 \\ & & 874/9^2 \end{bmatrix}.$$

Hence the sum

$$\sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(X_{ij} - \frac{X_{..}}{n_.}\right)\left(X_{ij} - \frac{X_{..}}{n_.}\right)' = \begin{bmatrix} 18 & -31/9 & -18/9 \\ & 1538/9^2 & 242/9^2 \\ & & 404/9^2 \end{bmatrix} + \begin{bmatrix} 8 & -6 & -6 \\ & 586/9^2 & 487/9^2 \\ & & 874/9^2 \end{bmatrix}$$

$$= \begin{bmatrix} 26 & -85/9 & -8 \\ -85/9 & 236/9 & 9 \\ -8 & 9 & 142/9 \end{bmatrix} = U + V$$

whose determinant is

$$26 \begin{vmatrix} 236/9 & 9 \\ 9 & 142/9 \end{vmatrix} + \frac{85}{9} \begin{vmatrix} -85/9 & 9 \\ -8 & 142/9 \end{vmatrix} - 8 \begin{vmatrix} -85/9 & 236/9 \\ -8 & 9 \end{vmatrix} = 8380.0549.$$

So, the observed values are as follows:

$$w = \frac{|V|}{|U+V|} = \frac{5416}{8380.055} = 0.6463$$

$$t_1 = \frac{1-w}{w} = \frac{0.3537}{0.6463} = 0.5413$$

$$\frac{n_.-1-p}{p} t_1 = \frac{5}{3} t_1 = \frac{5}{3}(0.5413) = 0.9022,$$

and $F_{p,n_.-1-p} = F_{3,5}$. Let us test the hypothesis at the 5% significance level. The critical value obtained from F-tables is $F_{3,5,0.05} = 9.01$. But since the observed value of F is $0.9022 < 9.01$, the hypothesis is not rejected. We expected this result because the hypothesis $A_1 = A_2 = A_3$ was not rejected. This example was mainly presented to illustrate the steps.

13.4. MANOVA for Two-Way Classification Data

As was done previously for the one-way classification, we will revisit the real scalar variable case first. Thus, we consider the case of two sets of treatments, instead of the single set analyzed in Sect. 13.3. In an agricultural experiment, suppose that we are considering r fertilizers as the first set of treatments, say $F_1, \ldots, F_r$, along with a set of s different varieties of corn, $V_1, \ldots, V_s$, as the second set of treatments. A randomized block experiment belongs to this category. In this case, r blocks of land, which are homogeneous with respect to all factors that may affect the yield of corn, such as precipitation, fertility of the soil, exposure to sunlight, drainage, and so on, are selected. Fertilizers $F_1, \ldots, F_r$ are applied to these r blocks at random, the first block receiving any one of $F_1, \ldots, F_r$, and so on. Each block is divided into s equivalent plots, all the plots being of the same size, shape, and so on. Then, the s varieties of corn are applied to each block at random, with one variety to each plot. Such an experiment is called a randomized block experiment. This experiment is then replicated t times. This replication is done so that possible interaction between fertilizers and varieties of corn could be tested. If the randomized block experiment is carried out only once, no interaction can be tested from such data because each plot will have only one observation. Interaction between the i-th fertilizer and j-th variety is a joint effect for the (F_i, V_j) combination, that is, the effect of F_i on the yield varies with the variety of corn. For instance, an interaction will be present if the effect of F_1 is different when combined with V_1 or V_2. In other words, there are individual effects and joint effects, a joint effect being referred to as an interaction between the two sets of treatments. As an example, consider one set of treatments consisting of r different methods of teaching and a second set of treatments that could be s levels of previous exposure of the students to the subject matter.

13.4.1. The model in a two-way classification

The additive, fixed effect, two-way classification or two-way layout model with inter-action is the following:

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \ i = 1, \ldots, r, \ j = 1, \ldots, s, \ k = 1, \ldots, t, \quad (13.4.1)$$

where μ is a general effect, α_i is the deviation from the general effect due to the i-th treatment of the first set, β_j is the deviation from the general effect due to the j-th treatment of the second set, and γ_{ij} is the effect due to interaction term or the joint effect of first and second sets of treatments. In a randomized block experiment, the treatments belonging to the first set are called "blocks" or "rows" and the treatments belonging to the second set are called "treatments" or "columns"; thus, the two sets correspond to rows, say $R_1, \ldots, R_r$, and columns, say $C_1, \ldots, C_s$. Then, γ_{ij} is the deviation from the general effect due to the combination (R_i, C_j). The random component e_{ijk} is the sum total contributions coming from all unknown factors and x_{ijk} is the observation resulting from the effect of the combination of treatments (R_i, C_j) at the k-th replication or k-th identical repetition of the experiment. In an agricultural setting, the observation may be the yield of corn whereas, in a teaching experiment, the observation may be the grade obtained by the "(i, j, k)"-th student. In a fixed effect model, all parameters $\mu, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s$ are assumed to be unknown constants. In a random effect model $\alpha_1, \ldots, \alpha_r$ or $\beta_1, \ldots, \beta_s$ or both sets are assumed to be random variables. We assume that $E[e_{ijk}] = 0$ and $\text{Var}(e_{ijk}) = \sigma^2 > 0$ for all i, j, k, where $E(\cdot)$ denotes the expected value of $(\cdot)$. In the present discussion, we will only consider the fixed effect model. Under this model, the data are called two-way classification data or two-way layout data because they can be classified according to the two sets of treatments, "rows" and "columns". Since we are not making any assumption about the distribution of e_{ijk}, and thereby that of x_{ijk}, we will apply the method of least squares to estimate the parameters.

13.4.2. Estimation of parameters in a two-way classification

The error sum of squares is

$$e_{ijk}^2 = \sum_{ijk} (x_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2.$$

Our first objective consists of isolating the sum of squares due to interaction and test the hypothesis of no interaction, that is, $H_o : \gamma_{ij} = 0$ for all i, j and k. If $\gamma_{ij} \neq 0$, part of the effect of the i-th row R_i is mixed up with the interaction and similarly, part of the effect of the j-th column, C_j, is intermingled with γ_{ij}, so that no hypothesis can be tested on the α_i's and β_j's unless γ_{ij} is zero or negligibly small or the hypothesis $\gamma_{ij} = 0$ is not

rejected. As well, on noting that in $[\mu + \alpha_i + \beta_j + \gamma_{ij}]$, the subscripts either appear none at a time, one at a time and both at a time, we may write $\mu + \alpha_i + \beta_j + \gamma_{ij} = m_{ij}$. Thus,

$$\sum_{ijk} e_{ijk}^2 = \sum_{ijk}(x_{ijk} - m_{ij})^2 \Rightarrow \frac{\partial}{\partial m_{ij}}[e_{ijk}^2] = 0$$

$$\Rightarrow \sum_k (x_{ijk} - m_{ij}) = 0 \Rightarrow x_{ij.} - t\,\hat{m}_{ij} = 0 \text{ or } \hat{m}_{ij} = \frac{x_{ij.}}{t}.$$

We employ the standard notation in this area, namely that a summation over a subscript is denoted by a dot. Then, the least squares minimum under the general model or the residual sum of squares, denoted by s^2, is given by

$$s^2 = \sum_{ijk}\left(x_{ijk} - \frac{x_{ij.}}{t}\right)^2. \tag{13.4.2}$$

Now, consider the hypothesis $H_o : \gamma_{ij} = 0$ for all i and j. Under this H_o, the model becomes

$$x_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \text{ or } \sum_{ijk} e_{ijk}^2 = \sum_{ijk}(x_{ijk} - \mu - \alpha_i - \beta_j)^2.$$

We differentiate this partially with respect to μ and α_i for a specific i, and to β_j for a specific j, and then equate the results to zero and solve to obtain estimates for μ, α_i and β_j. Since we have taken α_i, β_j and γ_{ij} as deviations from the general effect μ, we may let $\alpha_. = \alpha_1 + \cdots + \alpha_r = 0$, $\beta_. = \beta_1 + \cdots + \beta_s = 0$ and $\gamma_{i.} = 0$, for each i and $\gamma_{.j} = 0$ for each j, without any loss of generality. Then,

$$\frac{\partial}{\partial \mu}[e_{ijk}^2] = 0 \Rightarrow \left(\sum_{ijk} x_{ijk}\right) - rst\mu - st\alpha_. - rt\beta_. = 0 \Rightarrow \hat{\mu} = \frac{x_{...}}{rst}$$

$$\frac{\partial}{\partial \alpha_i}[e_{ijk}^2] = 0 \Rightarrow \sum_{jk}[x_{ijk} - \mu - \alpha_i - \beta_j] = 0$$

$$\Rightarrow x_{i..} - st\mu - st\alpha_i - t\beta_. = 0 \Rightarrow \hat{\alpha}_i = \frac{x_{i..}}{st} - \hat{\mu}$$

$$\frac{\partial}{\partial \beta_j}[e_{ijk}^2] = 0 \Rightarrow \sum_{ik}[x_{ijk} - \mu - \alpha_i - \beta_j] = 0$$

$$\Rightarrow x_{.j.} - rt\mu - t\alpha_. - rt\beta_j = 0 \Rightarrow \hat{\beta}_j = \frac{x_{.j.}}{rt} - \hat{\mu}\,.$$

Hence, the least squares minimum under the hypothesis H_o, denoted by s_0^2, is

$$
s_0^2 = \sum_{ijk} \left[\left(x_{ijk} - \frac{x_{...}}{rst} \right) - \left(\frac{x_{i..}}{st} - \frac{x_{...}}{rst} \right) - \left(\frac{x_{.j.}}{rt} - \frac{x_{...}}{rst} \right) \right]^2
$$

$$
= \sum_{ijk} \left(x_{ijk} - \frac{x_{...}}{rst} \right)^2 - st \sum_i \left(\frac{x_{i..}}{st} - \frac{x_{...}}{rst} \right)^2 - rt \sum_j \left(\frac{x_{.j.}}{rt} - \frac{x_{...}}{rst} \right)^2,
$$

the simplifications resulting from properties of summations with respect to subscripts. Thus, the sum of squares due to the hypothesis $H_o : \gamma_{ij} = 0$ for all i and j or the interaction sum of squares, denoted by s_γ^2 is the following:

$$
s_\gamma^2 = s_0^2 - s^2 = \sum_{ijk} \left(x_{ijk} - \frac{x_{...}}{rst} \right)^2 - st \sum_i \left(\frac{x_{i..}}{st} - \frac{x_{...}}{rst} \right)^2
$$

$$
- rt \sum_j \left(\frac{x_{.j.}}{rt} - \frac{x_{...}}{rst} \right)^2 - \sum_{ijk} \left(x_{ijk} - \frac{x_{ij.}}{t} \right)^2,
$$

and since

$$
\sum_{ijk} \left(x_{ijk} - \frac{x_{ij.}}{t} \right)^2 = \sum_{ijk} \left(x_{ijk} - \frac{x_{...}}{rst} \right)^2 - t \sum_{ij} \left(\frac{x_{ij.}}{t} - \frac{x_{...}}{rst} \right)^2,
$$

the sum of squares due to the hypothesis or attributable to the γ_{ij}'s, that is, due to interaction is

$$
s_\gamma^2 = t \sum_{ij} \left(\frac{x_{ij.}}{t} - \frac{x_{...}}{rst} \right)^2 - st \sum_i \left(\frac{x_{i..}}{st} - \frac{x_{...}}{rst} \right)^2 - rt \sum_j \left(\frac{x_{.j.}}{rt} - \frac{x_{...}}{rst} \right)^2. \quad (13.4.3)
$$

If the hypothesis $\gamma_{ij} = 0$ is not rejected, the effects of the γ_{ij}'s are deemed insignificant and then, setting the hypothesis $\gamma_{ij} = 0$, $\alpha_i = 0$, $i = 1, \ldots, r$, we obtain the sum of squares due to the α_i's or sum of squares due to the rows denoted as s_r^2, is

$$
s_r^2 = \sum_{ijk} \left(\frac{x_{i..}}{st} - \frac{x_{...}}{rst} \right)^2 = st \sum_{i=1}^r \left(\frac{x_{i..}}{st} - \frac{x_{...}}{rst} \right)^2. \quad (13.4.4)
$$

Similarly, the sum of squares attributable to the β_j's or due to the columns, denoted as s_c^2, is

$$
s_c^2 = rt \sum_{j=1}^s \left(\frac{x_{.j.}}{rt} - \frac{x_{...}}{rst} \right)^2. \quad (13.4.5)
$$

Observe that the sum of squares due to rows plus the sum of squares due to columns, once added to the interaction sum of squares, is the subtotal sum of squares, denoted by $s_{rc}^2 = t \sum_{ij} \left(\frac{x_{ij.}}{t} - \frac{x_{...}}{rst} \right)^2$ or this subtotal sum of squares is partitioned into the sum of squares due to the rows, due to the columns and due to interaction. This is equivalent to an ANOVA on the subtotals $\sum_k x_{ijk}$ or an ANOVA on a two-way classification with a single observation per cell. As has been pointed out, in that case, we cannot test for interaction, and moreover, this subtotal sum of squares plus the residual sum of squares is the grand total sum of squares. If we assume a normal distribution for the error terms, that is, $e_{ijk} \overset{iid}{\sim} N_1(0, \sigma^2), \ \sigma^2 > 0$, for all i, j, k, then under the hypothesis $H_o : \gamma_{ij} = 0$, it can be shown that

$$\frac{s_\gamma^2}{\sigma^2} \sim \chi_\nu^2, \quad \nu = (rs - 1) - (r - 1) - (s - 1) = (r - 1)(s - 1), \quad (13.4.6)$$

and the residual variation s^2 has the following distribution whether H_o holds or not:

$$\frac{s^2}{\sigma^2} \sim \chi_{\nu_1}^2, \quad \nu_1 = rst - 1 - (rs - 1) = rs(t - 1), \quad (13.4.7)$$

where s_γ^2 and s^2 are independently distributed. Then, under the hypothesis $\gamma_{ij} = 0$ for all i and j or when this hypothesis is not rejected, it can be established that

$$\frac{s_r^2}{\sigma^2} \sim \chi_{r-1}^2, \quad \frac{s_c^2}{\sigma^2} \sim \chi_{s-1}^2 \quad (13.4.8)$$

and s_r^2 and s^2 as well as s_c^2 and s^2 are independently distributed whenever $H_o : \gamma_{ij} = 0$ is not rejected. Hence, under the hypothesis,

$$\frac{s_\gamma^2/(r - 1)(s - 1)}{s^2/(rs(t - 1))} \sim F_{\nu, \nu_1}, \quad \nu = (r - 1)(s - 1), \ \nu_1 = rs(t - 1). \quad (13.4.9)$$

The total sum of squares is $\sum_{ijk} \left(x_{ijk} - \frac{x_{...}}{rst} \right)^2$. Thus, the first decomposition and the first part of ANOVA in this two-way classification scheme is the following:

Total variation $=$ Variation due to the subtotals $+$ Residual variation,

the second stage being

Variation due to the subtotals $=$ Variation due to the rows
$+$ Variation due to the columns $+$ Variation due to interaction,

and the resulting ANOVA table is the following:

ANOVA Table for the Two-Way Classification

Variation due to	df (1)	SS (2)	MS (3)=(2)/(1)
rows	$r-1$	$s_r^2 = st \sum_{i=1}^r (\frac{x_{i..}}{st} - \frac{x_{...}}{rst})^2$	$s_r^2/(r-1) = D_1$
columns	$s-1$	$s_c^2 = rt \sum_{j=1}^s (\frac{x_{.j.}}{rt} - \frac{x_{...}}{rst})^2$	$s_c^2/(s-1) = D_2$
interaction	$(r-1)(s-1)$	s_γ^2	$s_\gamma^2/(r-1)(s-1) = D_3$
subtotal	$rs-1$	$t \sum_{ij}(\frac{x_{ij.}}{t} - \frac{x_{...}}{rst})^2$	
residuals	$rs(t-1)$	s^2	$s^2/[rs(t-1)] = D$
total	$rst-1$	$\sum_{ijk}(x_{ijk} - \frac{x_{...}}{rst})^2$	

where df designates the number of degrees of freedom, SS means sum of squares, MS stands for mean squares, the expressions for the residual sum of squares is given in (13.4.2), that for the interaction in (13.4.3), that for the rows in (13.4.4) and that for columns in (13.4.5), respectively. Note that we test the hypothesis on the α_i's and β_j's or row effects and column effects, only if the hypothesis $\gamma_{ij} = 0$ is not rejected; otherwise there is no point in testing hypotheses on the α_i's and β_j's because they are confounded with the γ_{ij}'s.

13.5. Multivariate Extension of the Two-Way Layout

Instead of a single real scalar variable being studied, we consider a $p \times 1$ vector of real scalar variables. The multivariate two-way classification, the fixed effect model is the following:

$$X_{ijk} = M + A_i + B_j + \Gamma_{ij} + E_{ijk}, \tag{13.5.1}$$

for $i = 1, \ldots, r$, $j = 1, \ldots, s$, $k = 1, \ldots, t$, where M, A_i, B_j, Γ_{ij} and E_{ijk} are all $p \times 1$ vectors. In this case, M is a general effect, A_i is the deviation from the general effect due to the i-th row, B_j is the deviation from the general effect due to the j-th column, Γ_{ij} is the deviation from the general effect due to interaction between the rows and the columns and E_{ijk} is the vector of the random or error component. For convenience, the two sets of treatments are referred to as rows and columns, the first set as rows and the second, as columns. In a two-way layout, two sets of treatments are tested. As in the scalar case of Sect. 13.4, we can assume, without any loss of generality, that $\sum_i A_i = A_1 + \cdots + A_r = A_. = O$, $B_. = O$, $\sum_{i=1}^r \Gamma_{ij} = \Gamma_{.j} = O$ and $\sum_{j=1}^s \Gamma_{ij} = \Gamma_{i.} = O$. At this juncture, the procedures are parallel to those developed in Sect. 13.4 for the real scalar variable case. Instead of sums of squares, we now have sums of squares and cross products matrices. As

before, we may write $M_{ij} = M + A_i + B_j + \Gamma_{ij}$. Then, the trace of the sum of squares and cross products error matrix $E_{ijk}E'_{ijk}$ is minimized. Using the vector derivative operator, we have

$$\frac{\partial}{\partial M_{ij}} \text{tr}\left[\sum_{ijk} E_{ijk}E'_{ijk}\right] = O \Rightarrow \sum_{k}(X_{ijk} - M_{ij}) = O$$

$$\Rightarrow \hat{M}_{ij} = \frac{1}{t}X_{ij.},$$

so that the residual sum of squares and cross products matrix, denoted by S_{res}, is

$$S_{res} = \sum_{ijk}\left(X_{ijk} - \frac{X_{ij.}}{t}\right)\left(X_{ijk} - \frac{X_{ij.}}{t}\right)'. \tag{13.5.2}$$

All other derivations are analogous to those provided in the real scalar case. The sum of squares and cross products matrix due to interaction, denoted by S_{int} is the following:

$$S_{int} = t \sum_{ij}\left(\frac{X_{ij.}}{t} - \frac{X_{...}}{rst}\right)\left(\frac{X_{ij.}}{t} - \frac{X_{...}}{rst}\right)'$$

$$- st \sum_{i}\left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst}\right)\left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst}\right)'$$

$$- rt \sum_{j}\left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst}\right)\left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst}\right)'. \tag{13.5.3}$$

The sum of squares and cross products matrices due to the rows and columns are respectively given by

$$S_{row} = st \sum_{i=1}^{r}\left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst}\right)\left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst}\right)', \tag{13.5.4}$$

$$S_{col} = rt \sum_{j=1}^{s}\left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst}\right)\left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst}\right)'. \tag{13.5.5}$$

The sum of squares and cross products matrix for the subtotal is denoted by $S_{sub} = S_{row} + S_{col} + S_{int}$. The total sum of squares and cross products matrix, denoted by S_{tot}, is the following:

$$S_{tot} = \sum_{ijk}\left(X_{ijk} - \frac{X_{...}}{rst}\right)\left(X_{ijk} - \frac{X_{...}}{rst}\right)'. \tag{13.5.6}$$

We may now construct the MANOVA table. The following abbreviations are used: df stands for degrees of freedom of the corresponding Wishart matrix, SSP means the sum of squares and cross products matrix, MS stands for mean squares and is equal to SSP/df, and S_{row}, S_{col}, S_{res} and S_{tot} are respectively specified in (13.5.4), (13.5.5), (13.5.2) and (13.5.6).

MANOVA Table for a Two-Way Layout

	df	SSP	MS
Variation due to	(1)	(2)	(3)=(2)/(1)
rows	$r-1$	S_{row}	$S_{row}/(r-1)$
columns	$s-1$	S_{col}	$S_{col}/(s-1)$
interaction	$(r-1)(s-1)$	S_{int}	$S_{int}/[(r-1)(s-1)]$
subtotal	$rs-1$	S_{sub}	
residuals	$rs(t-1)$	S_{res}	$S_{res}/[rs(t-1)]$
total	$rst-1$	S_{tot}	

13.5.1. Likelihood ratio test for multivariate two-way layout

Under the assumption that the error or random components $E_{ijk} \overset{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$ for all i, j and k, the exponential part of the multivariate normal density excluding $-\frac{1}{2}$ is obtained as follows:

$$E'_{ijk}E_{ijk} = (X_{ijk} - M - A_i - B_j - \Gamma_{ij})'\Sigma^{-1}(X_{ijk} - M - A_i - B_j - \Gamma_{ij})$$
$$= \text{tr}[\Sigma^{-1}(X_{ijk} - M - A_i - B_j - \Gamma_{ij})(X_{ijk} - M - A_i - B_j - \Gamma_{ij})'] \Rightarrow$$
$$\sum_{ijk} E'_{ijk}E_{ijk} = \text{tr}\left\{\Sigma^{-1}\left[\sum_{ijk}(X_{ijk} - M - A_i - B_j - \Gamma_{ij})(X_{ijk} - M - A_i - B_j - \Gamma_{ij})'\right]\right\}.$$

Thus, the joint density of all the X_{ijk}'s, denoted by L, is

$$L = \frac{1}{(2\pi)^{\frac{prst}{2}}|\Sigma|^{\frac{rst}{2}}}$$
$$\times e^{-\frac{1}{2}\text{tr}[\Sigma^{-1}\sum_{ijk}(X_{ijk}-M-A_i-B_j-\Gamma_{ij})(X_{ijk}-M-A_i-B_j-\Gamma_{ij})']}.$$

The maximum likelihood estimates of M, A_i, B_j and Γ_{ij} are the same as the least squares estimates and hence, the maximum likelihood estimator (MLE) of Σ is the least squares

minimum which is the residual sum of squares and cross products matrix S or S_{res} (in the present notation), where

$$S_{res} = \sum_{ijk}(X_{ijk} - \hat{M} - \hat{A}_i - \hat{B}_j - \hat{\Gamma}_{ij})(X_{ijk} - \hat{M} - \hat{A}_i - \hat{B}_j - \hat{\Gamma}_{ij})'$$

$$= \sum_{ijk}\left(X_{ijk} - \frac{X_{ij.}}{t}\right)\left(X_{ijk} - \frac{X_{ij.}}{t}\right)'. \qquad (13.5.7)$$

This is the sample sum of squares and the cross products matrix under the general model and its determinant raised to the power of $\frac{rst}{2}$ is the quantity appearing in the numerator of the likelihood ratio criterion λ. Consider the hypothesis $H_o : \Gamma_{ij} = O$ for all i and j. Then, under this hypothesis, the estimator of Σ is S_0, where

$$S_0 = \sum_{ijk}\left[\left(X_{ijk} - \frac{X_{...}}{rst}\right) - \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst}\right) - \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst}\right)\right]$$

$$\times \left[\left(X_{ijk} - \frac{X_{...}}{rst}\right) - \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst}\right) - \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst}\right)\right]'$$

and $|S_0|^{\frac{rst}{2}}$ is the quantity appearing in the denominator of λ. However, $S_0 - S_{res} = S_{int}$ is the sum of squares and cross products matrix due to the interaction terms Γ_{ij}'s or to the hypothesis, so that $S_0 = S_{res} + S_{int}$. Therefore, λ is given by

$$\lambda = \frac{|S_{res}|^{\frac{rst}{2}}}{|S_{res} + S_{int}|^{\frac{rst}{2}}} \qquad (13.5.8)$$

Letting $w = \lambda^{\frac{2}{rst}}$,

$$w = \frac{|S_{res}|}{|S_{res} + S_{int}|}. \qquad (13.5.9)$$

It follows from results derived in Chap. 5 that $S_{res} \sim W_p(rs(t-1), \Sigma)$, $S_{int} \sim W_p((r-1)(s-1), \Sigma)$ under the hypothesis and S_{res} and S_{int} are independently distributed and hence, under H_o,

$$W = (S_{res} + S_{int})^{-\frac{1}{2}} S_{res}(S_{res} + S_{int})^{-\frac{1}{2}} \sim \text{real } p\text{-variate type-1 beta random variable}$$

with the parameters $(\frac{rs(t-1)}{2}, \frac{(r-1)(s-1)}{2})$. As well,

$$W_1 = S_{res}^{-\frac{1}{2}} S_{int} S_{res}^{-\frac{1}{2}} \sim \text{real p-variate type-2 beta random variable}$$

with the parameters $(\frac{(r-1)(s-1)}{2}, \frac{rs(t-1)}{2})$. Under H_o, the h-th arbitrary moments of w and λ, which are readily obtained from those of a real matrix-variate type-1 beta variable, are

$$E[w]^h = \left\{ \prod_{j=1}^{p} \frac{\Gamma(\nu_1 + \nu_2 - \frac{j-1}{2})}{\Gamma(\nu_1 - \frac{j-1}{2})} \right\} \left\{ \prod_{j=1}^{p} \frac{\Gamma(\nu_1 + h - \frac{j-1}{2})}{\Gamma(\nu_1 + \nu_2 + h - \frac{j-1}{2})} \right\} \qquad (13.5.10)$$

$$E[\lambda]^h = \left\{ \prod_{j=1}^{p} \frac{\Gamma(\nu_1 + \nu_2 - \frac{j-1}{2})}{\Gamma(\nu_1 - \frac{j-1}{2})} \right\} \left\{ \prod_{j=1}^{p} \frac{\Gamma(\nu_1 + \frac{rst}{2} h - \frac{j-1}{2})}{\Gamma(\nu_1 + \nu_2 + \frac{rst}{2} h - \frac{j-1}{2})} \right\} \qquad (13.5.11)$$

where $\nu_1 = \frac{rs(t-1)}{2}$ and $\nu_2 = \frac{(r-1)(s-1)}{2}$. Note that we reject the null hypothesis $H_o : \Gamma_{ij} = O, i = 1, \ldots, r, j = 1, \ldots, s$, for small values of w and λ. As explained in Sect. 13.3, the exact general density of w in (13.5.10) can be expressed in terms of a G-function and the exact general density of λ in (13.5.11) can be written in terms of a H-function. For the theory and applications of the G-function and the H-function, the reader may respectively refer to Mathai (1993) and Mathai et al. (2010).

13.5.2. Asymptotic distribution of λ in the MANOVA two-way layout

Consider the arbitrary h-th moment specified in (13.5.11). On expanding all the gamma functions for large values of rst in the constant part and for large values of $rst(1 + h)$ in the functional part by applying Stirling's formula or using the first term in the asymptotic expansion of a gamma function referring to (13.3.13), it can be verified that the h-th moment of λ behaves asymptotically as follows:

$$\lambda^h \to (1 + h)^{-\frac{p(r-1)(s-1)}{2}} \Rightarrow -2 \ln \lambda \to \chi^2_{p(r-1)(s-1)} \text{ as } rst \to \infty. \qquad (13.5.12)$$

Thus, for large values of rst, one can utilize this real scalar chisquare approximation for testing the hypothesis $H_o : \Gamma_{ij} = O$ for all i and j. We can work out a large number of exact distributions of w of (13.5.10) for special values of r, s, t, p. Observe that

$$E[w^h] = C \prod_{j=1}^{p} \frac{\Gamma(\frac{rs(t-1)}{2} - \frac{j-1}{2} + h)}{\Gamma(\frac{rs(t-1)}{2} + \frac{(r-1)(s-1)}{2} - \frac{j-1}{2} + h)} \qquad (13.5.13)$$

where C is the normalizing constant such that when $h = 0$, $E[w^h] = 1$. Thus, when $(r-1)(s-1)$ is a positive integer or when r or s is odd, the gamma functions cancel out, leaving a number of factors in the denominator which can be written as a sum by applying the partial fractions technique. For small values of p, the exact density will then be expressible as a sum involving only a few terms. For larger values of p, there will be repeated factors in the denominator, which complicates matters.

13.5.3. Exact densities of w in some special cases

We will consider several special cases of the h-th moment of w as given in (13.5.13).

Case (1): $p = 1$. In this case, h-th moment becomes

$$E[w^h] = C_1 \frac{\Gamma(\frac{rs(t-1)}{2} + h)}{\Gamma(\frac{rs(t-1)}{2} + \frac{(r-1)(s-1)}{2} + h)}$$

where C_1 is the associated normalizing constant. This is the h-th moment of a real scalar type-1 beta random variable with the parameters $(\frac{rs(t-1)}{2}, \frac{(r-1)(s-1)}{2})$. Hence $y = \frac{1-w}{w}$ is a real scalar type-2 beta with parameters $(\frac{(r-1)(s-1)}{2}, \frac{rs(t-1)}{2})$, and

$$\frac{rs(t-1)}{(r-1)(s-1)} y \sim F_{(r-1)(s-1), rs(t-1)}.$$

Accordingly, the test can be carried out by using this F-statistic. One would reject the null hypothesis $H_o : \Gamma_{ij} = O$ if the observed $F \geq F_{(r-1)(s-1), rs(t-1), \alpha}$ where $F_{(r-1)(s-1), rs(t-1), \alpha}$ is the upper $100\alpha\%$ percentile of this F-density. For example, for $r = 2, s = 3, t = 3$ and $\alpha = 0.05$, we have $F_{2,12,0.05} = 19.4$ from F-tables so that H_o would be rejected if the observed value of $F_{2,12} \geq 19.4$ at the specified significance level.

Case (2): $p = 2$. In this case, we have a ratio of two gamma functions differing by $\frac{1}{2}$. Combining the gamma functions in the numerator and in the denominator by using the duplication formula and proceeding as in Sect. 13.3 for the one-way layout, the statistic $t_1 = \frac{1-\sqrt{w}}{\sqrt{w}}$, and we have

$$\frac{rs(t-1)}{(r-1)(s-1)} t_1 \sim F_{2(r-1)(s-1), 2(rs(t-1)-1)},$$

so that the decision can be made as in Case (1).

Case (3): $(r-1)(s-1) = 1 \Rightarrow r = 2, s = 2$. In this case, all the gamma functions in (13.3.13) cancel out except the last one in the numerator and the first one in the denominator. This gamma ratio is that of a real scalar type-1 beta random variable with the parameters $(\frac{rs(t-1)+1-p}{2}, \frac{p}{2})$, and hence $y = \frac{1-w}{w}$ is a real scalar type-2 beta so that

$$\frac{rs(t-1)+1-p}{p} y \sim F_{p, rs(t-1)+1-p},$$

and decision can be made by making use of this F distribution as in Case (1).

Case(4): $(r-1)(s-1) = 2$. In this case,

$$E[w^h] = C_1 \prod_{j=1}^{p} \frac{1}{\frac{rs(t-1)}{2} - \frac{j-1}{2} + h}$$

with the corresponding normalizing constant C_1. This product of p factors can be expressed as a sum by using partial fractions. That is,

$$E[w^h] = C_1 \sum_{j=0}^{p-1} \frac{b_j}{a + h - \frac{j}{2}} \qquad (i)$$

where

$$b_j = \lim_{a+h \to \frac{j}{2}} [(a+h)(a+h-\tfrac{1}{2})\cdots(a+h-\tfrac{j-1}{2})(a+h-\tfrac{j+1}{2})\cdots(a+h-\tfrac{p-1}{2})],$$

$$\qquad (ii)$$

$$a = \frac{rs(t-1)}{2}.$$

Thus, the density of w, denoted by $f_w(w)$, which is available from *(i)* and *(ii)*, is the following:

$$f_w(w) = C_1 \sum_{j=0}^{p-1} b_j w^{a - \frac{j}{2} - 1}, \ 0 \le w \le 1,$$

and zero elsewhere. Some additional special cases could be worked out but the expressions would become complicated. For large values of rst, one can apply the asymptotic chisquare result given in (13.5.12) for testing the hypothesis $H_o : \Gamma_{ij} = O$.

Example 13.5.1.   An experiment is conducted among heart patients to stabilize their systolic pressure, diastolic pressure and heart rate or pulse around the standard numbers which are 120, 80 and 60, respectively. A random sample of 24 patients who may be considered homogeneous with respect to all factors of variation, such as age, weight group, race, gender, dietary habits, and so on, are selected. These 24 individuals are randomly divided into two groups of equal size. One group of 12 subjects are given the medication combination Med-1 and the other 12 are administered the medication combination Med-2. Then, the Med-1 group is randomly divided into three subgroups of 4 subjects. These subgroups are assigned exercise routines Ex-1, Ex-2, Ex-3. Similarly, the Med-2 group is also divided at random into 3 subgroups of 4 individuals who are respectively subjected to exercise routines Ex-1, Ex-2, Ex-3. After one week, the following observations are made $x_1 =$ current

reading on systolic pressure minus 120, $x_2 =$ current reading on diastolic pressure minus 80, $x_3 =$ current reading on heart rate minus 60. The structure of the two-way data layout is as follows:

	Ex-1	Ex-2	Ex-3
Med-1	four 3×1 vectors	four 3×1 vectors	four 3×1 vectors
Med-2	four 3×1 vectors	four 3×1 vectors	four 3×1 vectors

Let X_{ijk} be the k-th vector in the i-the row (i-th medication) and j-th column (j-th exercise routine). For convenience, the data are presented in matrix form:

$$A_{11} = [X_{111}, X_{112}, X_{113}, X_{114}], \quad A_{12} = [X_{121}, X_{122}, X_{123}, X_{124}],$$
$$A_{13} = [X_{131}, X_{132}, X_{133}, X_{134}], \quad A_{21} = [A_{211}, A_{212}, A_{213}, A_{214}],$$
$$A_{22} = [A_{221}, A_{222}, A_{223}, A_{224}], \quad A_{23} = [X_{231}, X_{232}, X_{233}, X_{234}];$$

$$A_{11} = \begin{bmatrix} -2 & 3 & 2 & 5 \\ 1 & -1 & 1 & -1 \\ 2 & -1 & -1 & 0 \end{bmatrix}, \ A_{12} = \begin{bmatrix} 1 & 4 & -1 & 4 \\ -2 & -2 & -3 & 3 \\ 3 & -2 & -1 & 0 \end{bmatrix}, \ A_{13} = \begin{bmatrix} 4 & -3 & 3 & 4 \\ 2 & -3 & 2 & 3 \\ 1 & -1 & 1 & -1 \end{bmatrix},$$

$$A_{21} = \begin{bmatrix} 2 & 0 & -2 & 0 \\ 1 & 4 & 1 & 2 \\ 2 & 1 & -1 & -2 \end{bmatrix}, \ A_{22} = \begin{bmatrix} 3 & -1 & -1 & 3 \\ 4 & 4 & 0 & 0 \\ 0 & 1 & -1 & 4 \end{bmatrix}, \ A_{23} = \begin{bmatrix} -2 & -1 & 0 & -1 \\ 1 & 4 & 0 & 3 \\ -2 & 0 & -2 & 0 \end{bmatrix}.$$

(1) Perform a two-way ANOVA on the first component, namely, x_1, the current reading minus 120; (2) Carry out a MANOVA on the full data.

Solution 13.5.1. We need the following quantities:

$$X_{11.} = \begin{bmatrix} 8 \\ 0 \\ 0 \end{bmatrix}, \ X_{12.} = \begin{bmatrix} 8 \\ -4 \\ 0 \end{bmatrix}, \ X_{13.} = \begin{bmatrix} 8 \\ 4 \\ 0 \end{bmatrix} \Rightarrow X_{1..} = \begin{bmatrix} 24 \\ 0 \\ 0 \end{bmatrix}$$

$$X_{21.} = \begin{bmatrix} 0 \\ 8 \\ 0 \end{bmatrix}, \ X_{22.} = \begin{bmatrix} 4 \\ 8 \\ 4 \end{bmatrix}, \ X_{23.} = \begin{bmatrix} -4 \\ 8 \\ -4 \end{bmatrix} \Rightarrow X_{2..} = \begin{bmatrix} 0 \\ 24 \\ 0 \end{bmatrix}$$

$$X_{.1.} = \begin{bmatrix} 8 \\ 8 \\ 0 \end{bmatrix}, \ X_{.2.} = \begin{bmatrix} 12 \\ 4 \\ 4 \end{bmatrix}, \ X_{.3.} = \begin{bmatrix} 4 \\ 12 \\ -4 \end{bmatrix} \Rightarrow X_{...} = \begin{bmatrix} 24 \\ 24 \\ 0 \end{bmatrix}$$

$$\bar{X} = \frac{X_{...}}{rst} = \frac{1}{24} \begin{bmatrix} 24 \\ 24 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \ r = 2, \ s = 3, \ t = 4.$$

By using the first elements in all these vectors, we will carry out a two-way ANOVA and answer the first question. Since these are all observations on real scalar variables, we will utilize lower-case letters to indicate scalar quantities. Thus, we have the following values:

$$s_{tot}^2 = \sum_{ij}(x_{ijk} - \bar{x})^2 = (-2 - 1)^2 + (3 - 1)^2 + \cdots + (-1 - 1)^2 = 136,$$

$$s_{row}^2 = 12\sum_{i=1}^{2}\left(\frac{x_{i..}}{12} - 1\right)^2 = 12[(2 - 1)^2 + (0 - 1)^2] = 24,$$

$$s_{col}^2 = 8\sum_{j=1}^{3}\left(\frac{x_{.j.}}{8} - 1\right)^2 = 8\left[(1 - 1)^2 + \left(\frac{3}{2} - 1\right)^2 + \left(\frac{1}{2} - 1\right)^2\right] = 4,$$

$$s_{int}^2 = 4\sum_{ij}\left(\frac{x_{ij.}}{4} - \frac{x_{i..}}{12} - \frac{x_{.j.}}{8} + 1\right)^2$$

$$= 4\left[\left(\frac{8}{4} - \frac{24}{12} - \frac{8}{8} + 1\right)^2 + \cdots + \left(-\frac{4}{4} - 0 - \frac{4}{8} + 1\right)^2\right] = 4,$$

$$s_{sub}^2 = 4\sum_{ij}\left(\frac{x_{ij.}}{4} - \frac{x_{...}}{24}\right)^2 = 4\left[\left(\frac{8}{4} - 1\right)^2 + \cdots + \left(-\frac{4}{4} - 1\right)\right] = 32,$$

$$s_{res}^2 = \sum_{ijk}\left(x_{ijk} - \frac{x_{ij.}}{4}\right)^2 = (-2 - 2)^2 + \cdots + (-1 + 1)^2 = 104,$$

$$s_{tot}^2 = \sum_{ijk}\left(x_{ijk} - \frac{x_{...}}{24}\right)^2 = (-2 - 1)^2 + (3 - 1)^2 + \cdots + (-1 - 1)^2 = 136.$$

All quantities have been calculated separately in order to verify the computations. We could have obtained the interaction sum of squares from the subtotal sum of squares minus the sum of squares due to rows and columns. Similarly, we could have obtained the residual sum of squares from the total sum of squares minus the subtotal sum of squares. We will set up the ANOVA table, where, as usual, df stands for degrees of freedom, SS means sum of squares and MS denotes mean squares:

ANOVA Table for a Two-Way Layout with Interaction

Variation due to	$df.$ (1)	SS (2)	MS (3)=(2)/(1)	F-ratio
rows	1	24	24	24/5.78
columns	2	4	2	2/5.78
interaction	2	4	2	2/5.78
subtotal	5	32		
residuals	18	104	5.78	
total	23	136		

For testing the hypothesis of no interaction, the F-value at the 5% significance level is $F_{2,18,0.05} \approx 19$. The observed value of this $F_{2,18}$ being $\frac{2}{5.78} \approx 0.35 < 19$, the hypothesis of no interaction is not rejected. Thus, we can test for the significance of the row and column effects. Consider the hypothesis $\alpha_1 = \alpha_2 = 0$. Then under this hypothesis and no interaction hypothesis, the F-ratio for the row sum of squares is $24/5.78 \approx 4.15 < 240 = F_{1,18,0.05}$, the tabulated value of $F_{1,18}$ at $\alpha = 0.05$. Therefore, this hypothesis is not rejected. Now, consider the hypothesis $\beta_1 = \beta_2 = \beta_3 = 0$. Since under this hypothesis and the hypothesis of no interaction, the F-ratio for the column sum of squares is $\frac{2}{5.78} = 0.35 < 19 = F_{2,18,0.05}$, it is not rejected either. Thus, the data show no significant interaction between exercise routine and medication, and no significant effect of the exercise routines or the two combinations of medications in bringing the systolic pressures closer to the standard value of 120.

We now carry out the computations needed to perform a MANOVA on the full data. We employ our standard notation by denoting vectors and matrices by capital letters. The sum of squares and cross products matrices for the rows and columns are the following, respectively denoted by S_{row} and S_{col} :

$$S_{row} = st \sum_{i=1}^{2} \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst} \right) \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst} \right)'$$

$$= 12 \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} [1, -1, 0] + \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} [-1, 1, 0] \right\} = 12 \begin{bmatrix} 24 & -24 & 0 \\ -24 & 24 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$S_{col} = rt \sum_{j=1}^{3} \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst} \right) \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst} \right)'$$

$$= 8 \left\{ O + \frac{1}{4} \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix} \right\} = \begin{bmatrix} 4 & -4 & 4 \\ -4 & 4 & -4 \\ 4 & -4 & 4 \end{bmatrix},$$

$$S_{int} = t \sum_{ij} \left(\frac{X_{ij.}}{t} - \frac{X_{i..}}{st} - \frac{X_{.j.}}{rt} + \frac{X_{...}}{rst} \right) \left(\frac{X_{ij.}}{t} - \frac{X_{i..}}{st} - \frac{X_{.j.}}{rt} + \frac{X_{...}}{rst} \right)'$$

$$= 4 \left\{ O + \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + O \right.$$

$$\left. + \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \right\} = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix},$$

$$S_{sub} = t \sum_{ij} \left(\frac{X_{ij.}}{t} - \bar{X} \right) \left(\frac{X_{ij.}}{t} - \bar{X} \right)'$$

$$= \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \cdots + \begin{bmatrix} 4 & -2 & 2 \\ -2 & 1 & -1 \\ 2 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 32 & -24 & 8 \\ -24 & 32 & 0 \\ 8 & 0 & 8 \end{bmatrix}.$$

We can verify the computations done so far as follows. The sum of squares and cross product matrices ought to be such that $S_{row} + S_{col} + S_{int} = S_{sub}$. These are

$$S_{row} + S_{col} + S_{int} = \begin{bmatrix} 24 & -24 & 0 \\ -24 & 24 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 4 & -4 & 4 \\ -4 & 4 & -4 \\ 4 & -4 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 32 & -24 & 8 \\ -24 & 32 & 0 \\ 8 & 0 & 8 \end{bmatrix} = S_{sub}.$$

Hence the result is verified. Now, the total and residual sums of squares and cross product matrices are

$$S_{tot} = \sum_{ijk}(X_{ijk} - \bar{X})(X_{ijk} - \bar{X})'$$

$$= \begin{bmatrix} 9 & 0 & -6 \\ 0 & 0 & 0 \\ -6 & 0 & 4 \end{bmatrix} + \cdots + \begin{bmatrix} 4 & -4 & 0 \\ -4 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 136 & 11 & 16 \\ 11 & 112 & 6 \\ 16 & 6 & 60 \end{bmatrix}$$

and

$$S_{res} = \sum_{ijk}\left(X_{ijk} - \frac{X_{ij.}}{t}\right)\left(X_{ijk} - \frac{X_{ij.}}{t}\right)'$$

$$= \begin{bmatrix} 16 & -14 & -8 \\ -4 & 1 & 2 \\ -8 & 2 & 4 \end{bmatrix} + \cdots + \begin{bmatrix} 2 & 3 & 0 \\ 3 & 10 & 2 \\ 0 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 104 & 35 & 8 \\ 35 & 80 & 6 \\ 8 & 6 & 52 \end{bmatrix}.$$

Then,

$$S_{res} + S_{int} = \begin{bmatrix} 104 & 35 & 8 \\ 35 & 80 & 6 \\ 8 & 6 & 52 \end{bmatrix} + \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix} = \begin{bmatrix} 108 & 39 & 12 \\ 39 & 84 & 10 \\ 12 & 10 & 56 \end{bmatrix}.$$

The above results are included in the following MANOVA table where df means degrees of freedom, SSP denotes a sum of squares and cross products matrix and MS is equal to SSP divided by the corresponding degrees of freedom:

MANOVA Table for a Two-Way Layout with Interaction

Variation due to	$df.$ (1)	SSP (2)	MS (3)=(2)/(1)
rows	1	S_{row}	S_{row}
columns	2	S_{col}	$\frac{1}{2}S_{col}$
interaction	2	S_{int}	$\frac{1}{2}S_{int}$
subtotal	5	S_{sub}	
residuals	18	S_{res}	$\frac{1}{18}S_{res}$
total	23	S_{tot}	

Then, the λ-criterion is

$$\lambda = \frac{|S_{res}|^{\frac{rst}{2}}}{|S_{res} + S_{int}|^{\frac{rst}{2}}} \implies w = \frac{|S_{res}|}{|S_{res} + S_{int}|}.$$

The determinants are as follows:

$$|S_{res}| = 104 \begin{vmatrix} 80 & 6 \\ 6 & 52 \end{vmatrix} - 35 \begin{vmatrix} 35 & 6 \\ 8 & 52 \end{vmatrix} + 8 \begin{vmatrix} 35 & 80 \\ 8 & 6 \end{vmatrix} = 363436$$

$$|S_{res} + S_{int}| = 108 \begin{vmatrix} 84 & 10 \\ 10 & 56 \end{vmatrix} - 39 \begin{vmatrix} 39 & 10 \\ 12 & 56 \end{vmatrix} + 12 \begin{vmatrix} 39 & 84 \\ 12 & 10 \end{vmatrix} = 409320.$$

Therefore,

$$w = \frac{363436}{409320} = 0.888 \implies \ln w = -0.118783 \implies -2\ln\lambda = 24(0.118783) = 2.8508.$$

We have explicit simple representations of the exact densities for the special cases $p = 1$, $p = 2$, $t = 2$, $t = 3$. However, our situation being $p = 3$, $t = 4$, they do not apply. A chisquare approximation is available for large values of rst, but our rst is only equal to 24. In this instance, $-2\ln\lambda \to \chi^2_{p(r-1)(s-1)} \simeq \chi^2_6$ as $rst \to \infty$. However, since the observed value of $-2\ln\lambda = 2.8508$ happens to be much smaller than the critical value resulting from the asymptotic distribution, which is $\chi^2_{6,0.05} = 12.59$ in this case, we can still safely decide not to reject the hypothesis $H_o : \Gamma_{ij} = O$ for all i and j, and go ahead and test for the main row and column effects, that is, the main effects of medical combinations Med-1 and Med-2 and the main effects of exercise routines Ex-1, Ex-2 and Ex-3. For testing the row effect, our hypothesis is $A_1 = A_2 = O$ and for testing the column effect, it is $B_1 = B_2 = B_3 = O$, given that $\Gamma_{ij} = O$ for all i and j. The corresponding likelihood ratio criteria are respectively,

$$\lambda_1 = \left(\frac{|S_{res}|}{|S_{res} + S_{row}|}\right)^{\frac{rst}{2}} \text{ and } \lambda_2 = \left(\frac{|S_{res}|}{|S_{res} + S_{col}|}\right)^{\frac{rst}{2}},$$

and we may utilize $w_j = \lambda_j^{\frac{2}{rst}}$, $j = 1, 2$. From previous calculations, we have

$$S_{res} + S_{row} = \begin{bmatrix} 104 & 35 & 8 \\ 35 & 80 & 6 \\ 8 & 6 & 52 \end{bmatrix} + \begin{bmatrix} 24 & -24 & 0 \\ -24 & 24 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 128 & 11 & 8 \\ 11 & 104 & 6 \\ 8 & 6 & 52 \end{bmatrix},$$

$$S_{res} + S_{col} = \begin{bmatrix} 104 & 35 & 8 \\ 35 & 80 & 6 \\ 8 & 6 & 52 \end{bmatrix} + \begin{bmatrix} 4 & -4 & 4 \\ -4 & 4 & -4 \\ 4 & -4 & 4 \end{bmatrix} = \begin{bmatrix} 108 & 31 & 12 \\ 31 & 84 & 2 \\ 12 & 2 & 56 \end{bmatrix}.$$

The required determinants are as follows:

$$|S_{res}| = 363436, \ |S_{res} + S_{row}| = 675724, \ |S_{res} + S_{col}| = 443176 \Rightarrow$$

$$w_1 = \frac{363436}{675724} = 0.5378468 \ \text{and} \ w_2 = \frac{363436}{443176} = 0.8200714,$$

$$-2\ln\lambda_1 = -24\ln 0.5378468 = 24(0.62018) = 14.88,$$

$$\chi^2_{p(r-1),\alpha} = \chi^2_{3,0.05} = 7.81 < 14.88; \tag{i}$$

$$-2\ln\lambda_2 = -24\ln 0.8200714 = 24(0.19836) = 4.76,$$

$$\chi^2_{p(s-1),\alpha} = \chi^2_{6,0.05} = 12.59 > 4.76. \tag{ii}$$

When $rst \to \infty$, $-2\ln\lambda_1 \to \chi^2_{p(r-1)}$ and $-2\ln\lambda_2 \to \chi^2_{p(s-1)}$, referring to Exercises 13.5.9 and 13.5.10, respectively. These results follow from the asymptotic expansion provided in Sect. 12.5.2. Even though $rst = 24$ is not that large, we may use these chisquare approximations for making decisions as the exact densities of w_1 and w_2 do not fall into the special cases previously discussed. When making use of the likelihood ratio criterion, we reject the hypotheses $A_1 = A_2 = O$ and $B_1 = B_2 = B_3 = O$ for small values of λ_1 and λ_2, respectively, which translates into large values of the approximate chisquare values. It is seen from *(i)* that the observed value $-2\ln\lambda_1$ is larger than the tabulated critical value and hence we reject the hypothesis $A_1 = A_2 = O$ at the 5% level. However, the hypothesis $B_1 = B_2 = B_3 = O$ is not rejected since the observed value is less than the critical value. We may conclude that the present data does not show any evidence of interaction between the exercise routines and medication combinations, that the exercise routine does not contribute significantly to bringing the subjects' initial readings closer to the standard values (120, 80, 60), whereas there is a possibility that the medical combinations Med-1 and Med-2 are effective in significantly causing the subjects' initial readings to approach standard values.

Note 13.5.1. It may be noticed from the MANOVA table that the second stage analysis will involve one observation per cell in a two-way layout, that is, the (i, j)-th cell will contain only one observation vector $X_{ij.}$ for the second stage analysis. Thus, $S_{sub} = S_{int} + S_{row} + S_{col}$ (the corresponding sum of squares in the real scalar case), and in this analysis with a single observation per cell, S_{int} acts as the residual sum of squares and cross products matrix (the residual sum of squares in the real scalar case). Accordingly, "interaction" cannot be tested when there is only a single observation per cell.

Exercises

13.1. In the ANOVA table obtained in Example 13.5.1, prove that (1) the sum of squares due to interaction and the residual sum of squares, (2) the sum of squares due to rows and residual sum of squares, (3) the sum of squares due to columns and residual sum of squares, are independently distributed under the normality assumption for the error variables, that is, $e_{ijk} \overset{iid}{\sim} N_1(0, \sigma^2), \ \sigma^2 > 0$.

13.2. In the MANOVA table obtained in Example 13.5.1, prove that (1) S_{int} and S_{res}, (2) S_{row} and S_{res}, (3) S_{col} and S_{res}, are independently distributed Wishart matrices when $E_{ijk} \overset{iid}{\sim} N_p(O, \Sigma), \ \Sigma > O$.

13.3. In a one-way layout, the following are the data on four treatments. (1) Carry out a complete ANOVA on the first component (including individual comparisons if the hypothesis of no interaction is not rejected). (2) Perform a full MANOVA on the full data.

$$\text{Treatment-1} \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \text{Treatment-2} \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

$$\text{Treatment-3} \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \text{Treatment-4} \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} -3 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 3 \end{bmatrix}.$$

13.4. Carry out a full one-way MANOVA on the following data:

$$\text{Treatment-1} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, \text{Treatment-2} \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 6 \\ 5 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix},$$

$$\text{Treatment-3} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}.$$

13.5. The following are the data on a two-way layout where A_{ij} denotes the data on the i-th row and j-th column cell. (1) Perform a complete ANOVA on the first component. (2) Carry out a full MANOVA on the full data. (3) Verify that $S_{row} + S_{col} + S_{int} = S_{sub}$ and $S_{sub} + S_{res} = S_{tot}$, (4) Evaluate the exact density of w.

$$A_{11} = \begin{matrix} 2 & 1 & 2 & -1 \\ 1 & 2 & 1 & 0 \\ -1 & 3 & 1 & 4 \end{matrix}, A_{12} = \begin{matrix} 1 & 3 & 1 & 1 \\ 4 & 4 & 1 & 1 \\ 6 & 3 & -1 & 0 \end{matrix}, A_{13} = \begin{matrix} 1 & -1 & 1 & -1 \\ -2 & 1 & 2 & 1 \\ 1 & -2 & 1 & -2 \end{matrix},$$

$$A_{21} = \begin{matrix} 3 & 4 & 2 & 3 \\ 2 & -1 & 2 & 3 \\ 3 & 5 & 2 & 2 \end{matrix}, A_{22} = \begin{matrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{matrix}, A_{23} = \begin{matrix} 2 & 3 & 3 & 0 \\ 3 & 2 & -1 & 2 \\ 3 & -3 & 2 & 4 \end{matrix}.$$

13.6. Carry out a complete MANOVA on the following data where A_{ij} indicates the data in the i-th row and j-th column cell.

$$A_{11} = \begin{matrix} 1 & -1 & 1 & 1 & 1 \\ 3 & 2 & 1 & 4 & 2 \end{matrix}, \quad A_{12} = \begin{matrix} 4 & 3 & 4 & 2 & 2 \\ 5 & 6 & 7 & 2 & 1 \end{matrix}, \quad A_{13} = \begin{matrix} 5 & 3 & 5 & 4 & 5 \\ 5 & 4 & 2 & 4 & 5 \end{matrix},$$

$$A_{21} = \begin{matrix} 0 & 1 & 1 & 0 & -2 \\ 2 & 2 & 0 & -1 & -1 \end{matrix}, \quad A_{22} = \begin{matrix} 6 & 7 & 6 & 4 & 5 \\ 4 & 5 & 2 & 3 & 4 \end{matrix}, \quad A_{23} = \begin{matrix} 1 & 0 & 1 & -1 & -1 \\ 1 & -1 & 2 & 1 & 2 \end{matrix}.$$

13.7. Under the hypothesis $A_1 = \cdots = A_r = O$, prove that $U_1 = (S_{res} + S_{row})^{-\frac{1}{2}} S_{res}$ $(S_{res} + S_{row})^{-\frac{1}{2}}$, is a real matrix-variate type-1 beta with the parameters $(\frac{rs(t-1)}{2}, \frac{r-1}{2})$ for $r \geq p$, $rs(t-1) \geq p$, when the hypothesis $\Gamma_{ij} = O$ for all i and j is not rejected or assuming that $\Gamma_{ij} = O$. The determinant of U_1 appears in the likelihood ratio criterion in this case.

13.8. Under the hypothesis $B_1 = \cdots = B_s = O$ when the hypothesis $\Gamma_{ij} = O$ is not rejected, or assuming that $\Gamma_{ij} = O$, prove that $U_2 = (S_{res} + S_{col})^{-\frac{1}{2}} S_{res} (S_{res} + S_{col})^{-\frac{1}{2}}$ is a real matrix-variate type-1 beta random variable with the parameters $(\frac{rs(t-1)}{2}, \frac{s-1}{2})$ for $s \geq p$, $rs(t-1) \geq p$. The determinant of U_2 appears in the likelihood ratio criterion for testing the main effect $B_j = O$, $j = 1, \ldots, s$.

13.9. Show that when $rst \to \infty$, $-2 \ln \lambda_1 \to \chi^2_{p(r-1)}$, that is, $-2 \ln \lambda_1$ asymptotically tends to a real scalar chisquare having $p(r-1)$ degrees of freedom, where $\lambda_1 = |U_1|$ and U_1 is as defined in Exercise 13.7. [Hint: Look into the general h-th moment of λ_1 in this case, which can be evaluated by using the density of U_1]. Hence for large values of rst, one can use this approximate chisquare distribution for testing the hypothesis $A_1 = \cdots = A_r = O$.

13.10. Show that when $rst \to \infty$, $-2 \ln \lambda_2 \to \chi^2_{p(s-1)}$, that is, $-2 \ln \lambda_2$ asymptotically converges to a real scalar chisquare having $p(s-1)$ degrees of freedom, where $\lambda_2 = |U_2|$ with U_2 as defined in Exercise 13.8. [Hint: Look at the h-th moment of λ_2]. For large values of rst, one can utilize this approximate chisquare distribution for testing the hypothesis $B_1 = \cdots = B_s = O$.

References

A.M. Mathai (1993): *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*, Oxford University Press, Oxford.

A.M. Mathai and H.J. Haubold (2017): *Probability and Statistics: A Course for Physicists and Engineers*, De Gruyter, Germany.

A.M. Mathai, R.K. Saxena and H.J. Haubold (2010): *The H-function: Theory and Applications*, Springer, New York.