Chapter 6

Comparisons of Several Multivariate Means

6.1 Introduction

The ideas developed in Chapter 5 can be extended to handle problems involving the comparison of several mean vectors. The theory is a little more complicated and rests on an assumption of multivariate normal distributions or large sample sizes. Similarly, the notation becomes a bit cumbersome. To circumvent these problems, we shall often review univariate procedures for comparing several means and then generalize to the corresponding multivariate cases by analogy. The numerical examples we present will help cement the concepts.

Because comparisons of means frequently (and should) emanate from designed experiments, we take the opportunity to discuss some of the tenets of good experimental practice. A *repeated measures* design, useful in behavioral studies, is explicitly considered, along with modifications required to analyze *growth curves*.

We begin by considering pairs of mean vectors. In later sections, we discuss several comparisons among mean vectors arranged according to treatment levels. The corresponding test statistics depend upon a partitioning of the total variation into pieces of variation attributable to the treatment sources and error. This partitioning is known as the *multivariate analysis of variance* (MANOVA).

6.2 Paired Comparisons and a Repeated Measures Design

Paired Comparisons

Measurements are often recorded under different sets of experimental conditions to see whether the responses differ significantly over these sets. For example, the efficacy of a new drug or of a saturation advertising campaign may be determined by comparing measurements before the "treatment" (drug or advertising) with those after the treatment. In other situations, *two or more* treatments can be administered to the same or similar experimental units, and responses can be compared to assess the effects of the treatments.

One rational approach to comparing two treatments, or the presence and absence of a single treatment, is to assign both treatments to the *same* or *identical* units (individuals, stores, plots of land, and so forth). The paired responses may then be analyzed by computing their differences, thereby eliminating much of the influence of extraneous unit-to-unit variation.

In the single response (univariate) case, let X_{j1} denote the response to treatment 1 (or the response before treatment), and let X_{j2} denote the response to treatment 2 (or the response after treatment) for the jth trial. That is, (X_{j1}, X_{j2}) are measurements recorded on the jth unit or jth pair of like units. By design, the n differences

$$D_j = X_{j1} - X_{j2}, \qquad j = 1, 2, \ldots, n \tag{6-1}$$

should reflect only the differential effects of the treatments.

Given that the differences D_j in (6-1) represent independent observations from an $N(\delta, \sigma_d^2)$ distribution, the variable

$$t = \frac{\bar{D} - \delta}{s_d/\sqrt{n}} \tag{6-2}$$

where

$$\bar{D} = \frac{1}{n}\sum_{j=1}^{n} D_j \quad \text{and} \quad s_d^2 = \frac{1}{n-1}\sum_{j=1}^{n}(D_j - \bar{D})^2 \tag{6-3}$$

has a t-distribution with $n - 1$ d.f. Consequently, an α-level test of

$$H_0: \delta = 0 \quad \text{(zero mean difference for treatments)}$$

versus

$$H_1: \delta \neq 0$$

may be conducted by comparing $|t|$ with $t_{n-1}(\alpha/2)$—the upper $100(\alpha/2)$th percentile of a t-distribution with $n - 1$ d.f. A $100(1 - \alpha)\%$ confidence interval for the mean difference $\delta = E(X_{j1} - X_{j2})$ is provided the statement

$$\bar{d} - t_{n-1}(\alpha/2)\frac{s_d}{\sqrt{n}} \leq \delta \leq \bar{d} + t_{n-1}(\alpha/2)\frac{s_d}{\sqrt{n}} \tag{6-4}$$

(For example, see [11].)

Additional notation is required for the multivariate extension of the paired-comparison procedure. It is necessary to distinguish between p responses, two treatments, and n experimental units. We label the p responses within the *j*th *unit* as

$$\begin{aligned}
X_{1j1} &= \text{variable 1 under treatment 1} \\
X_{1j2} &= \text{variable 2 under treatment 1} \\
&\vdots \\
X_{1jp} &= \text{variable } p \text{ under treatment 1} \\
\hline
X_{2j1} &= \text{variable 1 under treatment 2} \\
X_{2j2} &= \text{variable 2 under treatment 2} \\
&\vdots \\
X_{2jp} &= \text{variable } p \text{ under treatment 2}
\end{aligned}$$

and the p paired-difference random variables become

$$\begin{aligned} D_{j1} &= X_{1j1} - X_{2j1} \\ D_{j2} &= X_{1j2} - X_{2j2} \\ &\vdots \qquad \vdots \\ D_{jp} &= X_{1jp} - X_{2jp} \end{aligned} \tag{6-5}$$

Let $\mathbf{D}_j' = [D_{j1}, D_{j2}, \ldots, D_{jp}]$, and assume, for $j = 1, 2, \ldots, n$, that

$$E(\mathbf{D}_j) = \boldsymbol{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_p \end{bmatrix} \quad \text{and} \quad \text{Cov}(\mathbf{D}_j) = \boldsymbol{\Sigma}_d \tag{6-6}$$

If, in addition, $\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n$ are independent $N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d)$ random vectors, inferences about the vector of mean differences $\boldsymbol{\delta}$ can be based upon a T^2-statistic.

Specifically,

$$T^2 = n(\bar{\mathbf{D}} - \boldsymbol{\delta})'\mathbf{S}_d^{-1}(\bar{\mathbf{D}} - \boldsymbol{\delta}) \tag{6-7}$$

where

$$\bar{\mathbf{D}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{D}_j \quad \text{and} \quad \mathbf{S}_d = \frac{1}{n-1}\sum_{j=1}^{n} (\mathbf{D}_j - \bar{\mathbf{D}})(\mathbf{D}_j - \bar{\mathbf{D}})' \tag{6-8}$$

Result 6.1. Let the differences $\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n$ be a random sample from an $N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d)$ population. Then

$$T^2 = n(\bar{\mathbf{D}} - \boldsymbol{\delta})'\mathbf{S}_d^{-1}(\bar{\mathbf{D}} - \boldsymbol{\delta})$$

is distributed as an $[(n-1)p/(n-p)]F_{p,n-p}$ random variable, whatever the true $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}_d$.

If n and $n - p$ are both large, T^2 is approximately distributed as a χ_p^2 random variable, regardless of the form of the underlying population of differences.

Proof. The exact distribution of T^2 is a restatement of the summary in (5-6), with vectors of differences for the observation vectors. The approximate distribution of T^2, for n and $n - p$ large, follows from (4-28). ■

The condition $\boldsymbol{\delta} = \mathbf{0}$ is equivalent to "no average difference between the two treatments." For the ith variable, $\delta_i > 0$ implies that treatment 1 is larger, on average, than treatment 2. In general, inferences about $\boldsymbol{\delta}$ can be made using Result 6.1.

Given the observed differences $\mathbf{d}_j' = [d_{j1}, d_{j2}, \ldots, d_{jp}]$, $j = 1, 2, \ldots, n$, corresponding to the random variables in (6-5), an *α-level test* of H_0: $\boldsymbol{\delta} = \mathbf{0}$ *versus* H_1: $\boldsymbol{\delta} \neq \mathbf{0}$ for an $N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d)$ population rejects H_0 if the observed

$$T^2 = n\bar{\mathbf{d}}'\mathbf{S}_d^{-1}\bar{\mathbf{d}} > \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)$$

where $F_{p,n-p}(\alpha)$ is the upper (100α)th percentile of an F-distribution with p and $n - p$ d.f. Here $\bar{\mathbf{d}}$ and $\mathbf{S}_d$ are given by (6-8).

A $100(1-\alpha)\%$ *confidence region for* $\boldsymbol{\delta}$ consists of all $\boldsymbol{\delta}$ such that

$$(\bar{\mathbf{d}} - \boldsymbol{\delta})'\mathbf{S}_d^{-1}(\bar{\mathbf{d}} - \boldsymbol{\delta}) \leq \frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha) \tag{6-9}$$

Also, $100(1-\alpha)\%$ *simultaneous confidence intervals for the individual mean differences* δ_i are given by

$$\delta_i: \quad \bar{d}_i \pm \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{d_i}^2}{n}} \tag{6-10}$$

where $\bar{d}_i$ is the ith element of $\bar{\mathbf{d}}$ and $s_{d_i}^2$ is the ith diagonal element of $\mathbf{S}_d$.

For $n-p$ large, $[(n-1)p/(n-p)]F_{p,n-p}(\alpha) \doteq \chi_p^2(\alpha)$ and normality need not be assumed.

The *Bonferroni* $100(1-\alpha)\%$ *simultaneous confidence intervals* for the individual mean differences are

$$\delta_i: \quad \bar{d}_i \pm t_{n-1}\left(\frac{\alpha}{2p}\right) \sqrt{\frac{s_{d_i}^2}{n}} \tag{6-10a}$$

where $t_{n-1}(\alpha/2p)$ is the upper $100(\alpha/2p)$th percentile of a t-distribution with $n-1$ d.f.

Example 6.1 (Checking for a mean difference with paired observations) Municipal wastewater treatment plants are required by law to monitor their discharges into rivers and streams on a regular basis. Concern about the reliability of data from one of these self-monitoring programs led to a study in which samples of effluent were divided and sent to two laboratories for testing. One-half of each sample was sent to the Wisconsin State Laboratory of Hygiene, and one-half was sent to a private commercial laboratory routinely used in the monitoring program. Measurements of biochemical oxygen demand (BOD) and suspended solids (SS) were obtained, for $n = 11$ sample splits, from the two laboratories. The data are displayed in Table 6.1.

Table 6.1 Effluent Data

	Commercial lab		State lab of hygiene	
Sample j	x_{1j1} (BOD)	x_{1j2} (SS)	x_{2j1} (BOD)	x_{2j2} (SS)
1	6	27	25	15
2	6	23	28	13
3	18	64	36	22
4	8	44	35	29
5	11	30	15	31
6	34	75	44	64
7	28	26	42	30
8	71	124	54	64
9	43	54	34	56
10	33	30	29	20
11	20	14	39	21

Source: Data courtesy of S. Weber.

Do the two laboratories' chemical analyses agree? If differences exist, what is their nature?

The T^2-statistic for testing H_0: $\boldsymbol{\delta}' = [\delta_1, \delta_2] = [0, 0]$ is constructed from the differences of paired observations:

$d_{j1} = x_{1j1} - x_{2j1}$	−19	−22	−18	−27	−4	−10	−14	17	9	4	−19
$d_{j2} = x_{1j2} - x_{2j2}$	12	10	42	15	−1	11	−4	60	−2	10	−7

Here

$$\bar{\mathbf{d}} = \begin{bmatrix} \bar{d}_1 \\ \bar{d}_2 \end{bmatrix} = \begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix}, \qquad \mathbf{S}_d = \begin{bmatrix} 199.26 & 88.38 \\ 88.38 & 418.61 \end{bmatrix}$$

and

$$T^2 = 11[-9.36, \quad 13.27]\begin{bmatrix} .0055 & -.0012 \\ -.0012 & .0026 \end{bmatrix}\begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix} = 13.6$$

Taking $\alpha = .05$, we find that $[p(n-1)/(n-p)]F_{p,n-p}(.05) = [2(10)/9]F_{2,9}(.05) = 9.47$. Since $T^2 = 13.6 > 9.47$, we reject H_0 and conclude that there is a nonzero mean difference between the measurements of the two laboratories. It appears, from inspection of the data, that the commercial lab tends to produce lower BOD measurements and higher SS measurements than the State Lab of Hygiene. The 95% simultaneous confidence intervals for the mean differences δ_1 and δ_2 can be computed using (6-10). These intervals are

$$\delta_1: \bar{d}_1 \pm \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{d_1}^2}{n}} = -9.36 \pm \sqrt{9.47}\sqrt{\frac{199.26}{11}}$$

$$\text{or} \quad (-22.46, 3.74)$$

$$\delta_2: 13.27 \pm \sqrt{9.47}\sqrt{\frac{418.61}{11}} \quad \text{or} \quad (-5.71, 32.25)$$

The 95% *simultaneous confidence intervals* include zero, yet the hypothesis H_0: $\boldsymbol{\delta} = \mathbf{0}$ was rejected at the 5% level. What are we to conclude?

The evidence points toward real differences. The point $\boldsymbol{\delta} = \mathbf{0}$ falls outside the 95% *confidence region* for $\boldsymbol{\delta}$ (see Exercise 6.1), and this result is consistent with the T^2-test. The 95% simultaneous confidence coefficient applies to the *entire* set of intervals that could be constructed for all possible linear combinations of the form $a_1\delta_1 + a_2\delta_2$. The particular intervals corresponding to the choices $(a_1 = 1, a_2 = 0)$ and $(a_1 = 0, a_2 = 1)$ contain zero. Other choices of a_1 and a_2 will produce simultaneous intervals that do *not* contain zero. (If the hypothesis H_0: $\boldsymbol{\delta} = \mathbf{0}$ were not rejected, then *all* simultaneous intervals would include zero.)

The Bonferroni simultaneous intervals also cover zero. (See Exercise 6.2.)

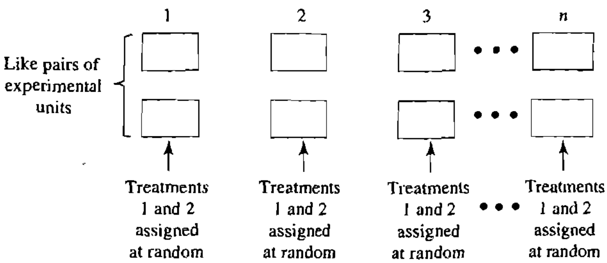
Our analysis assumed a normal distribution for the $\mathbf{D}_j$. In fact, the situation is further complicated by the presence of one or, possibly, two outliers. (See Exercise 6.3.) These data can be transformed to data more nearly normal, but with such a small sample, it is difficult to remove the effects of the outlier(s). (See Exercise 6.4.)

The numerical results of this example illustrate an unusual circumstance that can occur when making inferences. ■

The experimenter in Example 6.1 actually divided a sample by first shaking it and then pouring it rapidly back and forth into two bottles for chemical analysis. This was prudent because a simple division of the sample into two pieces obtained by pouring the top half into one bottle and the remainder into another bottle might result in more suspended solids in the lower half due to setting. The two laboratories would then not be working with the same, or even like, experimental units, and the conclusions would not pertain to laboratory competence, measuring techniques, and so forth.

Whenever an investigator can control the assignment of treatments to experimental units, an appropriate pairing of units and a randomized assignment of treatments can enhance the statistical analysis. Differences, if any, between supposedly identical units must be identified and most-alike units paired. Further, a random assignment of treatment 1 to one unit and treatment 2 to the other unit will help eliminate the systematic effects of uncontrolled sources of variation. Randomization can be implemented by flipping a coin to determine whether the first unit in a pair receives treatment 1 (heads) or treatment 2 (tails). The remaining treatment is then assigned to the other unit. A separate independent randomization is conducted for each pair. One can conceive of the process as follows:

Experimental Design for Paired Comparisons



We conclude our discussion of paired comparisons by noting that $\bar{\mathbf{d}}$ and $\mathbf{S}_d$, and hence T^2, may be calculated from the full-sample quantities $\bar{\mathbf{x}}$ and $\mathbf{S}$. Here $\bar{\mathbf{x}}$ is the $2p \times 1$ vector of sample averages for the p variables on the two treatments given by

$$\bar{\mathbf{x}}' = [\bar{x}_{11}, \bar{x}_{12}, \ldots, \bar{x}_{1p}, \bar{x}_{21}, \bar{x}_{22}, \ldots, \bar{x}_{2p}] \tag{6-11}$$

and $\mathbf{S}$ is the $2p \times 2p$ matrix of sample variances and covariances arranged as

$$\mathbf{S} = \begin{bmatrix} \underset{(p\times p)}{\mathbf{S}_{11}} & \underset{(p\times p)}{\mathbf{S}_{12}} \\ \underset{(p\times p)}{\mathbf{S}_{21}} & \underset{(p\times p)}{\mathbf{S}_{22}} \end{bmatrix} \tag{6-12}$$

The matrix $\mathbf{S}_{11}$ contains the sample variances and covariances for the p variables on treatment 1. Similarly, $\mathbf{S}_{22}$ contains the sample variances and covariances computed for the p variables on treatment 2. Finally, $\mathbf{S}_{12} = \mathbf{S}'_{21}$ are the matrices of sample covariances computed from observations on pairs of treatment 1 and treatment 2 variables.

Defining the matrix

$$\underset{(p\times 2p)}{\mathbf{C}} = \left[\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & -1 \end{array}\right] \tag{6-13}$$

$\uparrow$ $(p + 1)$st column

we can verify (see Exercise 6.9) that

$$\begin{aligned} \mathbf{d}_j &= \mathbf{C}\mathbf{x}_j, \qquad j = 1, 2, \ldots, n \\ \bar{\mathbf{d}} &= \mathbf{C}\bar{\mathbf{x}} \quad \text{and} \quad \mathbf{S}_d = \mathbf{C}\mathbf{S}\mathbf{C}' \end{aligned} \tag{6-14}$$

Thus,

$$T^2 = n\bar{\mathbf{x}}'\mathbf{C}'(\mathbf{C}\mathbf{S}\mathbf{C}')^{-1}\mathbf{C}\bar{\mathbf{x}} \tag{6-15}$$

and it is not necessary first to calculate the differences $\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n$. On the other hand, it is wise to calculate these differences in order to check normality and the assumption of a random sample.

Each row $\mathbf{c}'_i$ of the matrix $\mathbf{C}$ in (6-13) is a *contrast vector*, because its elements sum to zero. Attention is usually centered on contrasts when comparing treatments. Each contrast is perpendicular to the vector $\mathbf{1}' = [1, 1, \ldots, 1]$ since $\mathbf{c}'_i\mathbf{1} = 0$. The component $\mathbf{1}'\mathbf{x}_j$, representing the overall treatment sum, is ignored by the test statistic T^2 presented in this section.

A Repeated Measures Design for Comparing Treatments

Another generalization of the univariate paired t-statistic arises in situations where q treatments are compared with respect to a *single* response variable. Each subject or experimental unit receives each treatment once over successive periods of time. The jth observation is

$$\mathbf{X}_j = \begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jq} \end{bmatrix}, \qquad j = 1, 2, \ldots, n$$

where X_{ji} is the response to the ith treatment on the jth unit. The name *repeated measures* stems from the fact that all treatments are administered to each unit.

For comparative purposes, we consider contrasts of the components of $\boldsymbol{\mu} = E(\mathbf{X}_j)$. These could be

$$\begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \\ \vdots \\ \mu_1 - \mu_q \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{bmatrix} = \mathbf{C}_1 \boldsymbol{\mu}$$

or

$$\begin{bmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_2 \\ \vdots \\ \mu_q - \mu_{q-1} \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{bmatrix} = \mathbf{C}_2 \boldsymbol{\mu}$$

Both $\mathbf{C}_1$ and $\mathbf{C}_2$ are called *contrast matrices*, because their $q - 1$ rows are linearly independent and each is a contrast vector. The nature of the design eliminates much of the influence of unit-to-unit variation on treatment comparisons. Of course, the experimenter should randomize the order in which the treatments are presented to each subject.

When the treatment means are equal, $\mathbf{C}_1\boldsymbol{\mu} = \mathbf{C}_2\boldsymbol{\mu} = \mathbf{0}$. In general, the hypothesis that there are no differences in treatments (equal treatment means) becomes $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ for any choice of the contrast matrix $\mathbf{C}$.

Consequently, based on the contrasts $\mathbf{C}\mathbf{x}_j$ in the observations, we have means $\mathbf{C}\bar{\mathbf{x}}$ and covariance matrix $\mathbf{CSC}'$, and we test $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ using the T^2-statistic

$$T^2 = n(\mathbf{C}\bar{\mathbf{x}})'(\mathbf{CSC}')^{-1}\mathbf{C}\bar{\mathbf{x}}$$

Test for Equality of Treatments in a Repeated Measures Design

Consider an $N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population, and let $\mathbf{C}$ be a contrast matrix. An α-level test of H_0: $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ (equal treatment means) versus H_1: $\mathbf{C}\boldsymbol{\mu} \neq \mathbf{0}$ is as follows: Reject H_0 if

$$T^2 = n(\mathbf{C}\bar{\mathbf{x}})'(\mathbf{CSC}')^{-1}\mathbf{C}\bar{\mathbf{x}} > \frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,n-q+1}(\alpha) \qquad (6\text{-}16)$$

where $F_{q-1,n-q+1}(\alpha)$ is the upper (100α)th percentile of an F-distribution with $q - 1$ and $n - q + 1$ d.f. Here $\bar{\mathbf{x}}$ and $\mathbf{S}$ are the sample mean vector and covariance matrix defined, respectively, by

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j \quad \text{and} \quad \mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$$

It can be shown that T^2 does not depend on the particular choice of $\mathbf{C}$.[1]

[1] Any pair of contrast matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ must be related by $\mathbf{C}_1 = \mathbf{BC}_2$, with $\mathbf{B}$ nonsingular. This follows because each $\mathbf{C}$ has the largest possible number, $q - 1$, of linearly independent rows, all perpendicular to the vector $\mathbf{1}$. Then $(\mathbf{BC}_2)'(\mathbf{BC}_2\mathbf{S}\mathbf{C}_2'\mathbf{B}')^{-1}(\mathbf{BC}_2) = \mathbf{C}_2'\mathbf{B}'(\mathbf{B}')^{-1}(\mathbf{C}_2\mathbf{S}\mathbf{C}_2')^{-1}\mathbf{B}^{-1}\mathbf{BC}_2 = \mathbf{C}_2'(\mathbf{C}_2\mathbf{S}\mathbf{C}_2')^{-1}\mathbf{C}_2$, so T^2 computed with $\mathbf{C}_2$ or $\mathbf{C}_1 = \mathbf{BC}_2$ gives the same result.

A confidence region for contrasts $\mathbf{C}\boldsymbol{\mu}$, with $\boldsymbol{\mu}$ the mean of a normal population, is determined by the set of all $\mathbf{C}\boldsymbol{\mu}$ such that

$$n(\mathbf{C}\bar{\mathbf{x}} - \mathbf{C}\boldsymbol{\mu})'(\mathbf{CSC}')^{-1}(\mathbf{C}\bar{\mathbf{x}} - \mathbf{C}\boldsymbol{\mu}) \leq \frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,\,n-q+1}(\alpha) \tag{6-17}$$

where $\bar{\mathbf{x}}$ and $\mathbf{S}$ are as defined in (6-16). Consequently, simultaneous $100(1-\alpha)\%$ confidence intervals for single contrasts $\mathbf{c}'\boldsymbol{\mu}$ for any contrast vectors of interest are given by (see Result 5A.1)

$$\mathbf{c}'\boldsymbol{\mu}: \quad \mathbf{c}'\bar{\mathbf{x}} \pm \sqrt{\frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,n-q+1}(\alpha)} \sqrt{\frac{\mathbf{c}'\mathbf{S}\mathbf{c}}{n}} \tag{6-18}$$

Example 6.2 (Testing for equal treatments in a repeated measures design) Improved anesthetics are often developed by first studying their effects on animals. In one study, 19 dogs were initially given the drug pentobarbitol. Each dog was then administered carbon dioxide CO_2 at each of two pressure levels. Next, halothane (H) was added, and the administration of CO_2 was repeated. The response, milliseconds between heartbeats, was measured for the four treatment combinations:

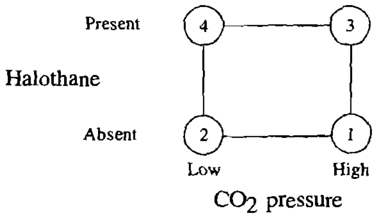


Table 6.2 contains the four measurements for each of the 19 dogs, where

Treatment 1 = high CO_2 pressure without H
Treatment 2 = low CO_2 pressure without H
Treatment 3 = high CO_2 pressure with H
Treatment 4 = low CO_2 pressure with H

We shall analyze the anesthetizing effects of CO_2 pressure and halothane from this repeated-measures design.

There are three treatment contrasts that might be of interest in the experiment. Let μ_1, μ_2, μ_3, and μ_4 correspond to the mean responses for treatments 1, 2, 3, and 4, respectively. Then

$$(\mu_3 + \mu_4) - (\mu_1 + \mu_2) = \begin{pmatrix} \text{Halothane contrast representing the} \\ \text{difference between the presence and} \\ \text{absence of halothane} \end{pmatrix}$$

$$(\mu_1 + \mu_3) - (\mu_2 + \mu_4) = \begin{pmatrix} CO_2 \text{ contrast representing the difference} \\ \text{between high and low } CO_2 \text{ pressure} \end{pmatrix}$$

$$(\mu_1 + \mu_4) - (\mu_2 + \mu_3) = \begin{pmatrix} \text{Contrast representing the influence} \\ \text{of halothane on } CO_2 \text{ pressure differences} \\ (\text{H}-CO_2 \text{ pressure "interaction"}) \end{pmatrix}$$

Table 6.2 Sleeping-Dog Data

	Treatment			
Dog	1	2	3	4
1	426	609	556	600
2	253	236	392	395
3	359	433	349	357
4	432	431	522	600
5	405	426	513	513
6	324	438	507	539
7	310	312	410	456
8	326	326	350	504
9	375	447	547	548
10	286	286	403	422
11	349	382	473	497
12	429	410	488	547
13	348	377	447	514
14	412	473	472	446
15	347	326	455	468
16	434	458	637	524
17	364	367	432	469
18	420	395	508	531
19	397	556	645	625

Source: Data courtesy of Dr. J. Atlee.

With $\boldsymbol{\mu}' = [\mu_1, \mu_2, \mu_3, \mu_4]$, the contrast matrix $\mathbf{C}$ is

$$\mathbf{C} = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

The data (see Table 6.2) give

$$\bar{\mathbf{x}} = \begin{bmatrix} 368.21 \\ 404.63 \\ 479.26 \\ 502.89 \end{bmatrix} \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} 2819.29 & & & \\ 3568.42 & 7963.14 & & \\ 2943.49 & 5303.98 & 6851.32 & \\ 2295.35 & 4065.44 & 4499.63 & 4878.99 \end{bmatrix}$$

It can be verified that

$$\mathbf{C}\bar{\mathbf{x}} = \begin{bmatrix} 209.31 \\ -60.05 \\ -12.79 \end{bmatrix}; \qquad \mathbf{CSC}' = \begin{bmatrix} 9432.32 & 1098.92 & 927.62 \\ 1098.92 & 5195.84 & 914.54 \\ 927.62 & 914.54 & 7557.44 \end{bmatrix}$$

and

$$T^2 = n(\mathbf{C}\bar{\mathbf{x}})'(\mathbf{CSC}')^{-1}(\mathbf{C}\bar{\mathbf{x}}) = 19(6.11) = 116$$

With $\alpha = .05$,

$$\frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,n-q+1}(\alpha) = \frac{18(3)}{16} F_{3,16}(.05) = \frac{18(3)}{16}(3.24) = 10.94$$

From (6-16), $T^2 = 116 > 10.94$, and we reject H_0: $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ (no treatment effects). To see which of the contrasts are responsible for the rejection of H_0, we construct 95% simultaneous confidence intervals for these contrasts. From (6-18), the contrast

$$\mathbf{c}_1'\boldsymbol{\mu} = (\mu_3 + \mu_4) - (\mu_1 + \mu_2) = \text{halothane influence}$$

is estimated by the interval

$$(\bar{x}_3 + \bar{x}_4) - (\bar{x}_1 + \bar{x}_2) \pm \sqrt{\frac{18(3)}{16} F_{3,16}(.05)} \sqrt{\frac{\mathbf{c}_1'\mathbf{S}\mathbf{c}_1}{19}} = 209.31 \pm \sqrt{10.94} \sqrt{\frac{9432.32}{19}}$$

$$= 209.31 \pm 73.70$$

where $\mathbf{c}_1'$ is the first row of $\mathbf{C}$. Similarly, the remaining contrasts are estimated by

CO_2 pressure influence $= (\mu_1 + \mu_3) - (\mu_2 + \mu_4)$:

$$-60.05 \pm \sqrt{10.94} \sqrt{\frac{5195.84}{19}} = -60.05 \pm 54.70$$

H–CO_2 pressure "interaction" $= (\mu_1 + \mu_4) - (\mu_2 + \mu_3)$:

$$-12.79 \pm \sqrt{10.94} \sqrt{\frac{7557.44}{19}} = -12.79 \pm 65.97$$

The first confidence interval implies that there is a halothane effect. The presence of halothane produces longer times between heartbeats. This occurs at both levels of CO_2 pressure, since the H–CO_2 pressure interaction contrast, $(\mu_1 + \mu_4) - (\mu_2 - \mu_3)$, is not significantly different from zero. (See the third confidence interval.) The second confidence interval indicates that there is an effect due to CO_2 pressure: The *lower* CO_2 pressure produces longer times between heartbeats.

Some caution must be exercised in our interpretation of the results because the trials with halothane must follow those without. The apparent H-effect may be due to a time trend. (Ideally, the time order of *all* treatments should be determined at random.) ■

The test in (6-16) is appropriate when the covariance matrix, $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$, cannot be assumed to have any special structure. If it is reasonable to assume that $\boldsymbol{\Sigma}$ has a particular structure, tests designed with this structure in mind have higher power than the one in (6-16). (For $\boldsymbol{\Sigma}$ with the equal correlation structure (8-14), see a discussion of the "randomized block" design in [17] or [22].)

6.3 Comparing Mean Vectors from Two Populations

A T^2-statistic for testing the equality of vector means from two multivariate populations can be developed by analogy with the univariate procedure. (See [11] for a discussion of the univariate case.) This T^2-statistic is appropriate for comparing responses from one set of experimental settings (population 1) with independent responses from another set of experimental settings (population 2). The comparison can be made without explicitly controlling for unit-to-unit variability, as in the paired-comparison case.

If possible, the experimental units should be randomly assigned to the sets of experimental conditions. Randomization will, to some extent, mitigate the effect of unit-to-unit variability in a subsequent comparison of treatments. Although some precision is lost relative to paired comparisons, the inferences in the two-population case are, ordinarily, applicable to a more general collection of experimental units simply because unit homogeneity is not required.

Consider a random sample of size n_1 from population 1 and a sample of size n_2 from population 2. The observations on p variables can be arranged as follows:

Sample	Summary statistics	
(Population 1) $\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1n_1}$	$\bar{\mathbf{x}}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} \mathbf{x}_{1j}$	$\mathbf{S}_1 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)'$
(Population 2) $\mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2n_2}$	$\bar{\mathbf{x}}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} \mathbf{x}_{2j}$	$\mathbf{S}_2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)'$

In this notation, the first subscript—1 or 2—denotes the population.

We want to make inferences about

$$\text{(mean vector of population 1)} - \text{(mean vector of population 2)} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2.$$

For instance, we shall want to answer the question, Is $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ (or, equivalently, is $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$)? Also, if $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq \mathbf{0}$, which component means are different?

With a few tentative assumptions, we are able to provide answers to these questions.

Assumptions Concerning the Structure of the Data

1. The sample $\mathbf{X}_{11}, \mathbf{X}_{12}, \ldots, \mathbf{X}_{1n_1}$, is a random sample of size n_1 from a p-variate population with mean vector $\boldsymbol{\mu}_1$ and covariance matrix $\boldsymbol{\Sigma}_1$.
2. The sample $\mathbf{X}_{21}, \mathbf{X}_{22}, \ldots, \mathbf{X}_{2n_2}$, is a random sample of size n_2 from a p-variate population with mean vector $\boldsymbol{\mu}_2$ and covariance matrix $\boldsymbol{\Sigma}_2$.
3. Also, $\mathbf{X}_{11}, \mathbf{X}_{12}, \ldots, \mathbf{X}_{1n_1}$, are independent of $\mathbf{X}_{21}, \mathbf{X}_{22}, \ldots, \mathbf{X}_{2n_2}$. (6-19)

We shall see later that, for large samples, this structure is sufficient for making inferences about the $p \times 1$ vector $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. However, when the sample sizes n_1 and n_2 are small, more assumptions are needed.

Further Assumptions When n_1 and n_2 Are Small

1. Both populations are multivariate normal.
2. Also, $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ (same covariance matrix). (6-20)

The second assumption, that $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, is much stronger than its univariate counterpart. Here we are assuming that several pairs of variances and covariances are nearly equal.

When $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, $\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)'$ is an estimate of $(n_1 - 1)\mathbf{\Sigma}$ and $\sum_{j=1}^{n_2}(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)'$ is an estimate of $(n_2 - 1)\mathbf{\Sigma}$. Consequently, we can pool the information in both samples in order to estimate the common covariance $\mathbf{\Sigma}$.

We set

$$\mathbf{S}_{\text{pooled}} = \frac{\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)' + \sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)'}{n_1 + n_2 - 2}$$
$$= \frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2 \tag{6-21}$$

Since $\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)'$ has $n_1 - 1$ d.f. and $\sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)'$ has $n_2 - 1$ d.f., the divisor $(n_1 - 1) + (n_2 - 1)$ in (6-21) is obtained by combining the two component degrees of freedom. [See (4-24).] Additional support for the pooling procedure comes from consideration of the multivariate normal likelihood. (See Exercise 6.11.)

To test the hypothesis that $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}_0$, a specified vector, we consider the squared statistical distance from $\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$ to $\boldsymbol{\delta}_0$. Now,

$$E(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) = E(\overline{\mathbf{X}}_1) - E(\overline{\mathbf{X}}_2) = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$$

Since the independence assumption in (6-19) implies that $\overline{\mathbf{X}}_1$ and $\overline{\mathbf{X}}_2$ are independent and thus $\text{Cov}(\overline{\mathbf{X}}_1, \overline{\mathbf{X}}_2) = \mathbf{0}$ (see Result 4.5), by (3-9), it follows that

$$\text{Cov}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) = \text{Cov}(\overline{\mathbf{X}}_1) + \text{Cov}(\overline{\mathbf{X}}_2) = \frac{1}{n_1}\mathbf{\Sigma} + \frac{1}{n_2}\mathbf{\Sigma} = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{\Sigma} \tag{6-22}$$

Because $\mathbf{S}_{\text{pooled}}$ estimates $\mathbf{\Sigma}$, we see that

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}$$

is an estimator of $\text{Cov}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)$.

The likelihood ratio test of

$$H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}_0$$

is based on the square of the statistical distance, T^2, and is given by (see [1]). Reject H_0 if

$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - \boldsymbol{\delta}_0)'\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\right]^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - \boldsymbol{\delta}_0) > c^2 \tag{6-23}$$

where the critical distance c^2 is determined from the distribution of the two-sample T^2-statistic.

Result 6.2. If $\mathbf{X}_{11}, \mathbf{X}_{12}, \ldots, \mathbf{X}_{1n_1}$ is a random sample of size n_1 from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\mathbf{X}_{21}, \mathbf{X}_{22}, \ldots, \mathbf{X}_{2n_2}$ is an independent random sample of size n_2 from $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, then

$$T^2 = [\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\right]^{-1} [\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]$$

is distributed as

$$\frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1+n_2-p-1}$$

Consequently,

$$P\left[(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\right]^{-1} (\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \le c^2\right] = 1 - \alpha \tag{6-24}$$

where

$$c^2 = \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1+n_2-p-1}(\alpha)$$

Proof. We first note that

$$\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 = \frac{1}{n_1}\mathbf{X}_{11} + \frac{1}{n_1}\mathbf{X}_{12} + \cdots + \frac{1}{n_1}\mathbf{X}_{1n_1} - \frac{1}{n_2}\mathbf{X}_{21} - \frac{1}{n_2}\mathbf{X}_{22} - \cdots - \frac{1}{n_2}\mathbf{X}_{2n_2}$$

is distributed as

$$N_p\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{\Sigma}\right)$$

by Result 4.8, with $c_1 = c_2 = \cdots = c_{n_1} = 1/n_1$ and $c_{n_1+1} = c_{n_1+2} = \cdots = c_{n_1+n_2} = -1/n_2$. According to (4-23),

$$(n_1 - 1)\mathbf{S}_1 \text{ is distributed as } W_{n_1-1}(\boldsymbol{\Sigma}) \text{ and } (n_2 - 1)\mathbf{S}_2 \text{ as } W_{n_2-1}(\boldsymbol{\Sigma})$$

By assumption, the $\mathbf{X}_{1j}$'s and the $\mathbf{X}_{2j}$'s are independent, so $(n_1 - 1)\mathbf{S}_1$ and $(n_2 - 1)\mathbf{S}_2$ are also independent. From (4-24), $(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2$ is then distributed as $\mathbf{W}_{n_1+n_2-2}(\boldsymbol{\Sigma})$. Therefore,

$$\begin{aligned}
T^2 &= \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} (\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))'\mathbf{S}_{\text{pooled}}^{-1} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} (\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \\
&= \begin{pmatrix}\text{multivariate normal} \\ \text{random vector}\end{pmatrix}' \left(\frac{\text{Wishart random matrix}}{\text{d.f.}}\right)^{-1} \begin{pmatrix}\text{multivariate normal} \\ \text{random vector}\end{pmatrix} \\
&= N_p(\mathbf{0}, \boldsymbol{\Sigma})' \left[\frac{W_{n_1+n_2-2}(\boldsymbol{\Sigma})}{n_1 + n_2 - 2}\right]^{-1} N_p(\mathbf{0}, \boldsymbol{\Sigma})
\end{aligned}$$

which is the T^2-distribution specified in (5-8), with n replaced by $n_1 + n_2 - 1$. [See (5-5) for the relation to F.] ∎

We are primarily interested in confidence regions for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. From (6-24), we conclude that all $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ within squared statistical distance c^2 of $\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$ constitute the confidence region. This region is an ellipsoid centered at the observed difference $\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$ and whose axes are determined by the eigenvalues and eigenvectors of $\mathbf{S}_{\text{pooled}}$ (or $\mathbf{S}_{\text{pooled}}^{-1}$).

Example 6.3 (Constructing a confidence region for the difference of two mean vectors) Fifty bars of soap are manufactured in each of two ways. Two characteristics, X_1 = lather and X_2 = mildness, are measured. The summary statistics for bars produced by methods 1 and 2 are

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 8.3 \\ 4.1 \end{bmatrix}, \qquad \mathbf{S}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}$$

$$\bar{\mathbf{x}}_2 = \begin{bmatrix} 10.2 \\ 3.9 \end{bmatrix}, \qquad \mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$$

Obtain a 95% confidence region for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.

We first note that $\mathbf{S}_1$ and $\mathbf{S}_2$ are approximately equal, so that it is reasonable to pool them. Hence, from (6-21),

$$\mathbf{S}_{\text{pooled}} = \frac{49}{98}\mathbf{S}_1 + \frac{49}{98}\mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}$$

Also,

$$\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 = \begin{bmatrix} -1.9 \\ .2 \end{bmatrix}$$

so the confidence ellipse is centered at $[-1.9, .2]'$. The eigenvalues and eigenvectors of $\mathbf{S}_{\text{pooled}}$ are obtained from the equation

$$0 = |\mathbf{S}_{\text{pooled}} - \lambda\mathbf{I}| = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 5-\lambda \end{vmatrix} = \lambda^2 - 7\lambda + 9$$

so $\lambda = (7 \pm \sqrt{49 - 36})/2$. Consequently, $\lambda_1 = 5.303$ and $\lambda_2 = 1.697$, and the corresponding eigenvectors, $\mathbf{e}_1$ and $\mathbf{e}_2$, determined from

$$\mathbf{S}_{\text{pooled}}\mathbf{e}_i = \lambda_i\mathbf{e}_i, \qquad i = 1, 2$$

are

$$\mathbf{e}_1 = \begin{bmatrix} .290 \\ .957 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_2 = \begin{bmatrix} .957 \\ -.290 \end{bmatrix}$$

By Result 6.2,

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right)c^2 = \left(\frac{1}{50} + \frac{1}{50}\right)\frac{(98)(2)}{(97)}F_{2,97}(.05) = .25$$

since $F_{2,97}(.05) = 3.1$. The confidence ellipse extends

$$\sqrt{\lambda_i}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)c^2} = \sqrt{\lambda_i}\sqrt{.25}$$

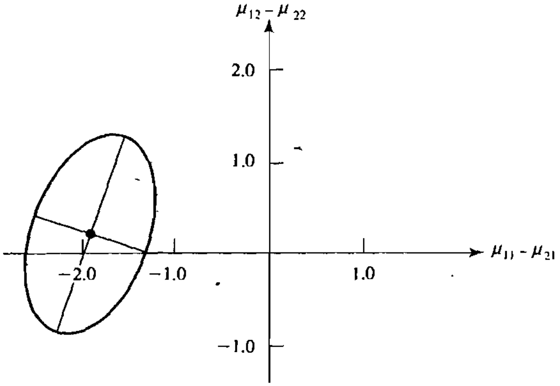


Figure 6.1 95% confidence ellipse for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.

units along the eigenvector $\mathbf{e}_i$, or 1.15 units in the $\mathbf{e}_1$ direction and .65 units in the $\mathbf{e}_2$ direction. The 95% confidence ellipse is shown in Figure 6.1. Clearly, $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ is not in the ellipse, and we conclude that the two methods of manufacturing soap produce different results. It appears as if the two processes produce bars of soap with about the same mildness (X_2), but those from the second process have more lather (X_1). ■

Simultaneous Confidence Intervals

It is possible to derive simultaneous confidence intervals for the components of the vector $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. These confidence intervals are developed from a consideration of all possible linear combinations of the differences in the mean vectors. It is assumed that the parent multivariate populations are normal with a common covariance $\boldsymbol{\Sigma}$.

Result 6.3. Let $c^2 = [(n_1 + n_2 - 2)p/(n_1 + n_2 - p - 1)]F_{p,n_1+n_2-p-1}(\alpha)$. With probability $1 - \alpha$.

$$\mathbf{a}'(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) \pm c\sqrt{\mathbf{a}'\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\mathbf{a}}$$

will cover $\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ for all $\mathbf{a}$. In particular $\mu_{1i} - \mu_{2i}$ will be covered by

$$(\overline{X}_{1i} - \overline{X}_{2i}) \pm c\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s_{ii,\text{pooled}}} \qquad \text{for } i = 1, 2, \ldots, p$$

Proof. Consider univariate linear combinations of the observations

$$\mathbf{X}_{11}, \mathbf{X}_{12}, \ldots, \mathbf{X}_{1n_1} \quad \text{and} \quad \mathbf{X}_{21}, \mathbf{X}_{22}, \ldots, \mathbf{X}_{2n_2}$$

given by $\mathbf{a}'\mathbf{X}_{1j} = a_1X_{1j1} + a_2X_{1j2} + \cdots + a_pX_{1jp}$ and $\mathbf{a}'\mathbf{X}_{2j} = a_1X_{2j1} + a_2X_{2j2} + \cdots + a_pX_{2jp}$. These linear combinations have sample means and covariances $\mathbf{a}'\overline{\mathbf{X}}_1$, $\mathbf{a}'\mathbf{S}_1\mathbf{a}$ and $\mathbf{a}'\overline{\mathbf{X}}_2$, $\mathbf{a}'\mathbf{S}_2\mathbf{a}$, respectively, where $\overline{\mathbf{X}}_1$, $\mathbf{S}_1$, and $\overline{\mathbf{X}}_2$, $\mathbf{S}_2$ are the mean and covariance statistics for the two original samples. (See Result 3.5.) When both parent populations have the same covariance matrix, $s^2_{1,\mathbf{a}} = \mathbf{a}'\mathbf{S}_1\mathbf{a}$ and $s^2_{2,\mathbf{a}} = \mathbf{a}'\mathbf{S}_2\mathbf{a}$

are both estimators of $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$, the common population variance of the linear combinations $\mathbf{a}'\mathbf{X}_1$ and $\mathbf{a}'\mathbf{X}_2$. Pooling these estimators, we obtain

$$\begin{aligned} s^2_{\mathbf{a},\text{pooled}} &= \frac{(n_1 - 1)s^2_{1,\mathbf{a}} + (n_2 - 1)s^2_{2,\mathbf{a}}}{(n_1 + n_2 - 2)} \\ &= \mathbf{a}'\left[\frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2\right]\mathbf{a} \qquad (6\text{-}25) \\ &= \mathbf{a}'\mathbf{S}_{\text{pooled}}\mathbf{a} \end{aligned}$$

To test $H_0\colon \mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mathbf{a}'\boldsymbol{\delta}_0$, on the basis of the $\mathbf{a}'\mathbf{X}_{1j}$ and $\mathbf{a}'\mathbf{X}_{2j}$, we can form the square of the univariate two-sample t-statistic

$$t^2_{\mathbf{a}} = \frac{[\mathbf{a}'(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) - \mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s^2_{\mathbf{a},\text{pooled}}} = \frac{[\mathbf{a}'(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))]^2}{\mathbf{a}'\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\mathbf{a}} \qquad (6\text{-}26)$$

According to the maximization lemma with $\mathbf{d} = (\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))$ and $\mathbf{B} = (1/n_1 + 1/n_2)\mathbf{S}_{\text{pooled}}$ in (2-50),

$$\begin{aligned} t^2_{\mathbf{a}} &\le (\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))'\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\right]^{-1}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \\ &= T^2 \end{aligned}$$

for all $\mathbf{a} \neq \mathbf{0}$. Thus,

$$\begin{aligned} (1 - \alpha) &= P[T^2 \le c^2] = P[t^2_{\mathbf{a}} \le c^2, \quad \text{for all } \mathbf{a}] \\ &= P\left[\left|\mathbf{a}'(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) - \mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right| \le c\sqrt{\mathbf{a}'\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\mathbf{a}} \quad \text{for all } \mathbf{a}\right] \end{aligned}$$

where c^2 is selected according to Result 6.2. ■

Remark. For testing $H_0\colon \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$, the linear combination $\hat{\mathbf{a}}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$, with coefficient vector $\hat{\mathbf{a}} \propto \mathbf{S}^{-1}_{\text{pooled}}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$, quantifies the largest population difference. That is, if T^2 rejects H_0, then $\hat{\mathbf{a}}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ will have a nonzero mean. Frequently, we try to interpret the components of this linear combination for both subject matter and statistical importance.

Example 6.4 (Calculating simultaneous confidence intervals for the differences in mean components) Samples of sizes $n_1 = 45$ and $n_2 = 55$ were taken of Wisconsin homeowners with and without air conditioning, respectively. (Data courtesy of Statistical Laboratory, University of Wisconsin.) Two measurements of electrical usage (in kilowatt hours) were considered. The first is a measure of total *on*-peak consumption (X_1) during July, and the second is a measure of total *off*-peak consumption (X_2) during July. The resulting summary statistics are

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix}, \qquad \mathbf{S}_1 = \begin{bmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix}, \qquad n_1 = 45$$

$$\bar{\mathbf{x}}_2 = \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix}, \qquad \mathbf{S}_2 = \begin{bmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix}, \qquad n_2 = 55$$

(The off-peak consumption is higher than the on-peak consumption because there are more off-peak hours in a month.)

Let us find 95% simultaneous confidence intervals for the differences in the mean components.

Although there appears to be somewhat of a discrepancy in the sample variances, for illustrative purposes we proceed to a calculation of the pooled sample covariance matrix. Here

$$\mathbf{S}_{\text{pooled}} = \frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2 = \begin{bmatrix} 10963.7 & 21505.5 \\ 21505.5 & 63661.3 \end{bmatrix}$$

and

$$c^2 = \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p,n_1+n_2-p-1}(\alpha) = \frac{98(2)}{97} F_{2,97}(.05)$$

$$= (2.02)(3.1) = 6.26$$

With $\boldsymbol{\mu}_1' - \boldsymbol{\mu}_2' = [\mu_{11} - \mu_{21}, \mu_{12} - \mu_{22}]$, the 95% simultaneous confidence intervals for the population differences are

$$\mu_{11} - \mu_{21}: (204.4 - 130.0) \pm \sqrt{6.26}\sqrt{\left(\frac{1}{45} + \frac{1}{55}\right)10963.7}$$

or

$$21.7 \le \mu_{11} - \mu_{21} \le 127.1 \qquad (\textit{on}\text{-peak})$$

$$\mu_{12} - \mu_{22}: (556.6 - 355.0) \pm \sqrt{6.26}\sqrt{\left(\frac{1}{45} + \frac{1}{55}\right)63661.3}$$

or

$$74.7 \le \mu_{12} - \mu_{22} \le 328.5 \qquad (\textit{off}\text{-peak})$$

We conclude that there is a difference in electrical consumption between those with air-conditioning and those without. This difference is evident in both on-peak and off-peak consumption.

The 95% confidence ellipse for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ is determined from the eigenvalue-eigenvector pairs $\lambda_1 = 71323.5$, $\mathbf{e}_1' = [.336, .942]$ and $\lambda_2 = 3301.5$, $\mathbf{e}_2' = [.942, -.336]$.

Since

$$\sqrt{\lambda_1}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)c^2} = \sqrt{71323.5}\sqrt{\left(\frac{1}{45} + \frac{1}{55}\right)6.26} = 134.3$$

and

$$\sqrt{\lambda_2}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)c^2} = \sqrt{3301.5}\sqrt{\left(\frac{1}{45} + \frac{1}{55}\right)6.26} = 28.9$$

we obtain the 95% confidence ellipse for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ sketched in Figure 6.2 on page 291. Because the confidence ellipse for the difference in means does not cover $\mathbf{0}' = [0, 0]$, the T^2-statistic will reject $H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ at the 5% level.

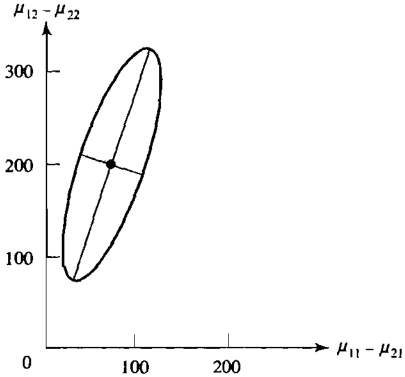


Figure 6.2 95% confidence ellipse for $\boldsymbol{\mu}_1' - \boldsymbol{\mu}_2' = (\mu_{11} - \mu_{21}, \mu_{12} - \mu_{22})$.

The coefficient vector for the linear combination most responsible for rejection is proportional to $\mathbf{S}_{\text{pooled}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$. (See Exercise 6.7.) ■

The *Bonferroni* $100(1 - \alpha)\%$ *simultaneous confidence intervals* for the p population mean differences are

$$\mu_{1i} - \mu_{2i}: \quad (\bar{x}_{1i} - \bar{x}_{2i}) \pm t_{n_1+n_2-2}\left(\frac{\alpha}{2p}\right)\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s_{ii,\text{pooled}}}$$

where $t_{n_1+n_2-2}(\alpha/2p)$ is the upper $100(\alpha/2p)$th percentile of a t-distribution with $n_1 + n_2 - 2$ d.f.

The Two-Sample Situation When $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$

When $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$. we are unable to find a "distance" measure like T^2, whose distribution does not depend on the unknowns $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. Bartlett's test [3] is used to test the equality of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ in terms of generalized variances. Unfortunately, the conclusions can be seriously misleading when the populations are nonnormal. Nonnormality and unequal covariances cannot be separated with Bartlett's test. (See also Section 6.6.) A method of testing the equality of two covariance matrices that is less sensitive to the assumption of multivariate normality has been proposed by Tiku and Balakrishnan [23]. However, more practical experience is needed with this test before we can recommend it unconditionally.

We suggest, without much factual support, that any discrepancy of the order $\sigma_{1,ii} = 4\sigma_{2,ii}$, or vice versa, is probably serious. This is true in the univariate case. The size of the discrepancies that are critical in the multivariate situation probably depends, to a large extent, on the number of variables p.

A transformation may improve things when the marginal variances are quite different. However, for n_1 and n_2 large, we can avoid the complexities due to unequal covariance matrices.

Result 6.4. Let the sample sizes be such that $n_1 - p$ and $n_2 - p$ are large. Then, an approximate $100(1 - \alpha)\%$ confidence ellipsoid for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ is given by all $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ satisfying

$$[\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]' \left[\frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 \right]^{-1} [\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \leq \chi_p^2(\alpha)$$

where $\chi_p^2(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with p d.f. Also, $100(1 - \alpha)\%$ simultaneous confidence intervals for all linear combinations $\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ are provided by

$$\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad \text{belongs to} \quad \mathbf{a}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \pm \sqrt{\chi_p^2(\alpha)} \sqrt{\mathbf{a}' \left(\frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 \right) \mathbf{a}}$$

Proof. From (6-22) and (3-9),

$$E(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$$

and

$$\text{Cov}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) = \text{Cov}(\overline{\mathbf{X}}_1) + \text{Cov}(\overline{\mathbf{X}}_2) = \frac{1}{n_1} \boldsymbol{\Sigma}_1 + \frac{1}{n_2} \boldsymbol{\Sigma}_2$$

By the central limit theorem, $\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2$ is nearly $N_p[\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, n_1^{-1}\boldsymbol{\Sigma}_1 + n_2^{-1}\boldsymbol{\Sigma}_2]$. If $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ were known, the square of the statistical distance from $\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2$ to $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ would be

$$[\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]' \left(\frac{1}{n_1} \boldsymbol{\Sigma}_1 + \frac{1}{n_2} \boldsymbol{\Sigma}_2 \right)^{-1} [\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]$$

This squared distance has an approximate χ_p^2-distribution, by Result 4.7. When n_1 and n_2 are large, with high probability, $\mathbf{S}_1$ will be close to $\boldsymbol{\Sigma}_1$ and $\mathbf{S}_2$ will be close to $\boldsymbol{\Sigma}_2$. Consequently, the approximation holds with $\mathbf{S}_1$ and $\mathbf{S}_2$ in place of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively.

The results concerning the simultaneous confidence intervals follow from Result 5 A.1. ■

Remark. If $n_1 = n_2 = n$, then $(n - 1)/(n + n - 2) = 1/2$, so

$$\frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 = \frac{1}{n} (\mathbf{S}_1 + \mathbf{S}_2) = \frac{(n - 1) \mathbf{S}_1 + (n - 1) \mathbf{S}_2}{n + n - 2} \left(\frac{1}{n} + \frac{1}{n} \right)$$

$$= \mathbf{S}_{\text{pooled}} \left(\frac{1}{n} + \frac{1}{n} \right)$$

With equal sample sizes, the large sample procedure is essentially the same as the procedure based on the pooled covariance matrix. (See Result 6.2.) In one dimension, it is well known that the effect of unequal variances is least when $n_1 = n_2$ and greatest when n_1 is much less than n_2 or vice versa.

Example 6.5 (Large sample procedures for inferences about the difference in means) We shall analyze the electrical-consumption data discussed in Example 6.4 using the large sample approach. We first calculate

$$\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2 = \frac{1}{45}\begin{bmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix} + \frac{1}{55}\begin{bmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix}$$

$$= \begin{bmatrix} 464.17 & 886.08 \\ 886.08 & 2642.15 \end{bmatrix}$$

The 95% simultaneous confidence intervals for the linear combinations

$$\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = [1,0]\begin{bmatrix} \mu_{11} - \mu_{21} \\ \mu_{12} - \mu_{22} \end{bmatrix} = \mu_{11} - \mu_{21}$$

and

$$\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = [0,1]\begin{bmatrix} \mu_{11} - \mu_{21} \\ \mu_{12} - \mu_{22} \end{bmatrix} = \mu_{12} - \mu_{22}$$

are (see Result 6.4)

$$\mu_{11} - \mu_{21}: \quad 74.4 \pm \sqrt{5.99}\,\sqrt{464.17} \quad \text{or} \quad (21.7, 127.1)$$

$$\mu_{12} - \mu_{22}: \quad 201.6 \pm \sqrt{5.99}\,\sqrt{2642.15} \quad \text{or} \quad (75.8, 327.4)$$

Notice that these intervals differ negligibly from the intervals in Example 6.4, where the pooling procedure was employed. The T^2-statistic for testing $H_0\colon \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ is

$$T^2 = [\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2]'\left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1}[\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2]$$

$$= \begin{bmatrix} 204.4 - 130.0 \\ 556.6 - 355.0 \end{bmatrix}'\begin{bmatrix} 464.17 & 886.08 \\ 886.08 & 2642.15 \end{bmatrix}^{-1}\begin{bmatrix} 204.4 - 130.0 \\ 556.6 - 355.0 \end{bmatrix}$$

$$= [74.4 \quad 201.6](10^{-4})\begin{bmatrix} 59.874 & -20.080 \\ -20.080 & 10.519 \end{bmatrix}\begin{bmatrix} 74.4 \\ 201.6 \end{bmatrix} = 15.66$$

For $\alpha = .05$, the critical value is $\chi_2^2(.05) = 5.99$ and, since $T^2 = 15.66 > \chi_2^2(.05) = 5.99$, we reject H_0.

The most critical linear combination leading to the rejection of H_0 has coefficient vector

$$\hat{\mathbf{a}} \propto \left(\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right)^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = (10^{-4})\begin{bmatrix} 59.874 & -20.080 \\ -20.080 & 10.519 \end{bmatrix}\begin{bmatrix} 74.4 \\ 201.6 \end{bmatrix}$$

$$= \begin{bmatrix} .041 \\ .063 \end{bmatrix}$$

The difference in *off*-peak electrical consumption between those with air conditioning and those without contributes more than the corresponding difference in *on*-peak consumption to the rejection of $H_0\colon \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$. ■

A statistic similar to T^2 that is less sensitive to outlying observations for small and moderately sized samples has been developed by Tiku and Singh [24]. However, if the sample size is moderate to large, Hotelling's T^2 is remarkably unaffected by slight departures from normality and/or the presence of a few outliers.

An Approximation to the Distribution of T^2 for Normal Populations When Sample Sizes Are Not Large

One can test H_0: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ when the population covariance matrices are unequal even if the two sample sizes are not large, provided the two populations are multivariate normal. This situation is often called the multivariate Behrens-Fisher problem. The result requires that both sample sizes n_1 and n_2 are greater than p, the number of variables. The approach depends on an approximation to the distribution of the statistic

$$T^2 = (\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))'\left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \qquad (6\text{-}27)$$

which is identical to the large sample statistic in Result 6.4. However, instead of using the chi-square approximation to obtain the critical value for testing H_0 the recommended approximation for smaller samples (see [15] and [19]) is given by

$$T^2 = \frac{\nu p}{\nu - p + 1}F_{p,\nu-p+1} \qquad (6\text{-}28)$$

where the degrees of freedom ν are estimated from the sample covariance matrices using the relation

$$\nu = \frac{p + p^2}{\sum_{i=1}^{2}\frac{1}{n_i}\left\{\text{tr}\left[\left(\frac{1}{n_i}\mathbf{S}_i\left(\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right)^{-1}\right)^2\right] + \left(\text{tr}\left[\frac{1}{n_i}\mathbf{S}_i\left(\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right)^{-1}\right]\right)^2\right\}} \qquad (6\text{-}29)$$

where $\min(n_1, n_2) \le \nu \le n_1 + n_2$. This approximation reduces to the usual Welch solution to the Behrens-Fisher problem in the univariate ($p = 1$) case.

With moderate sample sizes and two normal populations, the approximate level α test for equality of means rejects H_0: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ if

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))'\left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) > \frac{\nu p}{\nu - p + 1}F_{p,\nu-p+1}(\alpha)$$

where the degrees of freedom ν are given by (6-29). This procedure is consistent with the large samples procedure in Result 6.4 except that the critical value $\chi_p^2(\alpha)$ is replaced by the larger constant $\frac{\nu p}{\nu - p + 1}F_{p,\nu-p+1}(\alpha)$.

Similarly, the approximate $100(1 - \alpha)\%$ confidence region is given by all $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ such that

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))'\left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \le \frac{\nu p}{\nu - p + 1}F_{p,\nu-p+1}(\alpha) \qquad (6\text{-}30)$$

For normal populations, the approximation to the distribution of T^2 given by (6-28) and (6-29) usually gives reasonable results.

Example 6.6 (The approximate T^2 distribution when $\Sigma_1 \neq \Sigma_2$) Although the sample sizes are rather large for the electrical consumption data in Example 6.4, we use these data and the calculations in Example 6.5 to illustrate the computations leading to the approximate distribution of T^2 when the population covariance matrices are unequal.

We first calculate

$$\frac{1}{n_1}\mathbf{S}_1 = \frac{1}{45}\begin{bmatrix} 13825.2 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix} = \begin{bmatrix} 307.227 & 529.409 \\ 529.409 & 1624.609 \end{bmatrix}$$

$$\frac{1}{n_2}\mathbf{S}_2 = \frac{1}{55}\begin{bmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix} = \begin{bmatrix} 156.945 & 356.667 \\ 356.667 & 1017.536 \end{bmatrix}$$

and using a result from Example 6.5,

$$\left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1} = (10^{-4})\begin{bmatrix} 59.874 & -20.080 \\ -20.080 & 10.519 \end{bmatrix}$$

Consequently,

$$\frac{1}{n_1}\mathbf{S}_1\left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1} =$$

$$\begin{bmatrix} 307.227 & 529.409 \\ 529.409 & 1624.609 \end{bmatrix}(10^{-4})\begin{bmatrix} 59.874 & -20.080 \\ -20.080 & 10.519 \end{bmatrix} = \begin{bmatrix} .776 & -.060 \\ -.092 & .646 \end{bmatrix}$$

and

$$\left(\frac{1}{n_1}\mathbf{S}_1\left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1}\right)^2 = \begin{bmatrix} .776 & -.060 \\ -.092 & .646 \end{bmatrix}\begin{bmatrix} .776 & -.060 \\ -.092 & .646 \end{bmatrix} = \begin{bmatrix} .608 & -.085 \\ -.131 & .423 \end{bmatrix}$$

Further,

$$\frac{1}{n_2}\mathbf{S}_2\left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1} =$$

$$\begin{bmatrix} 156.945 & 356.667 \\ 356.667 & 1017.536 \end{bmatrix}(10^{-4})\begin{bmatrix} 59.874 & -20.080 \\ -20.080 & 10.519 \end{bmatrix} = \begin{bmatrix} .224 & -.060 \\ .092 & .354 \end{bmatrix}$$

and

$$\left(\frac{1}{n_2}\mathbf{S}_2\left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1}\right)^2 = \begin{bmatrix} .224 & .060 \\ -.092 & .354 \end{bmatrix}\begin{bmatrix} .224 & .060 \\ -.092 & .354 \end{bmatrix} = \begin{bmatrix} .055 & .035 \\ .053 & .131 \end{bmatrix}$$

Then

$$\frac{1}{n_1}\left\{\mathrm{tr}\left[\left(\frac{1}{n_1}\mathbf{S}_1\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)^{-1}\right)^2\right]+\left(\mathrm{tr}\left[\frac{1}{n_1}\mathbf{S}_1\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)^{-1}\right]\right)^2\right\}$$

$$=\frac{1}{45}\{(.608+.423)+(.776+.646)^2\}=.0678$$

$$\frac{1}{n_2}\left\{\mathrm{tr}\left[\left(\frac{1}{n_2}\mathbf{S}_2\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)^{-1}\right)^2\right]+\left(\mathrm{tr}\left[\frac{1}{n_2}\mathbf{S}_2\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)^{-1}\right]\right)^2\right\}$$

$$=\frac{1}{55}\{(.055+.131)+(.224+.354)^2\}=.0095$$

Using (6-29), the estimated degrees of freedom ν is

$$\nu=\frac{2+2^2}{.0678+.0095}=77.6$$

and the $\alpha = .05$ critical value is

$$\frac{\nu p}{\nu-p+1}F_{p,\nu-p+1}(.05)=\frac{77.6\times 2}{77.6-2+1}F_{2,77.6-2+1}(.05)=\frac{155.2}{76.6}3.12=6.32$$

From Example 6.5, the observed value of the test statistic is $T^2 = 15.66$ so the hypothesis $H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ is rejected at the 5% level. This is the same conclusion reached with the large sample procedure described in Example 6.5. ■

As was the case in Example 6.6, the $F_{p,\nu-p+1}$ distribution can be defined with noninteger degrees of freedom. A slightly more conservative approach is to use the integer part of ν.

6.4 Comparing Several Multivariate Population Means (One-Way MANOVA)

Often, more than two populations need to be compared. Random samples, collected from each of g populations, are arranged as

$$\begin{array}{ll}\text{Population 1:} & \mathbf{X}_{11},\mathbf{X}_{12},\ldots,\mathbf{X}_{1n_1}\\ \text{Population 2:} & \mathbf{X}_{21},\mathbf{X}_{22},\ldots,\mathbf{X}_{2n_2}\\ \vdots & \vdots\\ \text{Population } g\text{:} & \mathbf{X}_{g1},\mathbf{X}_{g2},\ldots,\mathbf{X}_{gn_g}\end{array}\qquad(6\text{-}31)$$

MANOVA is used first to investigate whether the population mean vectors are the same and, if not, which mean components differ significantly.

Assumptions about the Structure of the Data for One-Way MANOVA

1. $\mathbf{X}_{\ell 1},\mathbf{X}_{\ell 2},\ldots,\mathbf{X}_{\ell n_\ell}$, is a random sample of size n_ℓ from a population with mean $\boldsymbol{\mu}_\ell$, $\ell = 1, 2, \ldots, g$. The random samples from different populations are independent.

2. All populations have a common covariance matrix $\boldsymbol{\Sigma}$.
3. Each population is multivariate normal.

Condition 3 can be relaxed by appealing to the central limit theorem (Result 4.13) when the sample sizes n_ℓ are large.

A review of the univariate analysis of variance (ANOVA) will facilitate our discussion of the multivariate assumptions and solution methods.

A Summary of Univariate ANOVA

In the univariate situation, the assumptions are that $X_{\ell 1}, X_{\ell 2}, \ldots, X_{\ell n_\ell}$ is a random sample from an $N(\mu_\ell, \sigma^2)$ population, $\ell = 1, 2, \ldots, g$, and that the random samples are independent. Although the null hypothesis of equality of means could be formulated as $\mu_1 = \mu_2 = \cdots = \mu_g$, it is customary to regard μ_ℓ as the sum of an overall mean component, such as μ, and a component due to the specific population. For instance, we can write $\mu_\ell = \mu + (\mu_\ell - \mu)$ or $\mu_\ell = \mu + \tau_\ell$ where $\tau_\ell = \mu_\ell - \mu$.

Populations usually correspond to different sets of experimental conditions, and therefore, it is convenient to investigate the deviations τ_ℓ associated with the ℓth population (treatment).

The *reparameterization*

$$\begin{array}{ccccc} \mu_\ell & = & \mu & + & \tau_\ell \\ \begin{pmatrix} \ell\text{th population} \\ \text{mean} \end{pmatrix} & & \begin{pmatrix} \text{overall} \\ \text{mean} \end{pmatrix} & & \begin{pmatrix} \ell\text{th population} \\ \text{(treatment) effect} \end{pmatrix} \end{array} \tag{6-32}$$

leads to a restatement of the hypothesis of equality of means. The null hypothesis becomes

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_g = 0$$

The response $X_{\ell j}$, distributed as $N(\mu + \tau_\ell, \sigma^2)$, can be expressed in the suggestive form

$$\begin{array}{cccccc} X_{\ell j} = & \mu & + & \tau_\ell & + & e_{\ell j} \\ & \text{(overall mean)} & & \begin{pmatrix} \text{treatment} \\ \text{effect} \end{pmatrix} & & \begin{pmatrix} \text{random} \\ \text{error} \end{pmatrix} \end{array} \tag{6-33}$$

where the $e_{\ell j}$ are independent $N(0, \sigma^2)$ random variables. To define uniquely the model parameters and their least squares estimates, it is customary to impose the constraint $\sum_{\ell=1}^{g} n_\ell \tau_\ell = 0$.

Motivated by the decomposition in (6-33), the analysis of variance is based upon an analogous decomposition of the observations,

$$\begin{array}{ccccccc} x_{\ell j} & = & \bar{x} & + & (\bar{x}_\ell - \bar{x}) & + & (x_{\ell j} - \bar{x}_\ell) \\ \text{(observation)} & & \begin{pmatrix} \text{overall} \\ \text{sample mean} \end{pmatrix} & & \begin{pmatrix} \text{estimated} \\ \text{treatment effect} \end{pmatrix} & & \text{(residual)} \end{array} \tag{6-34}$$

where $\bar{x}$ is an estimate of μ, $\hat{\tau}_\ell = (\bar{x}_\ell - \bar{x})$ is an estimate of τ_ℓ, and $(x_{\ell j} - \bar{x}_\ell)$ is an estimate of the error $e_{\ell j}$.

Example 6.7 (The sum of squares decomposition for univariate ANOVA) Consider the following independent samples.

Population 1: 9, 6, 9

Population 2: 0, 2

Population 3: 3, 1, 2

Since, for example, $\bar{x}_3 = (3 + 1 + 2)/3 = 2$ and $\bar{x} = (9 + 6 + 9 + 0 + 2 + 3 + 1 + 2)/8 = 4$, we find that

$$\begin{aligned} 3 = x_{31} &= \bar{x} + (\bar{x}_3 - \bar{x}) + (x_{31} - \bar{x}_3) \\ &= 4 + (2 - 4) + (3 - 2) \\ &= 4 + (-2) + 1 \end{aligned}$$

Repeating this operation for each observation, we obtain the arrays

$$\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & \\ 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 & \\ 4 & 4 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 & \\ -2 & -2 & -2 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & \\ 1 & -1 & 0 \end{pmatrix}$$

$$\underset{(x_{\ell j})}{\text{observation}} = \underset{(\bar{x})}{\text{mean}} + \underset{(\bar{x}_\ell - \bar{x})}{\text{treatment effect}} + \underset{(x_{\ell j} - \bar{x}_\ell)}{\text{residual}}$$

The question of equality of means is answered by assessing whether the contribution of the treatment array is large relative to the residuals. (Our estimates $\hat{\tau}_\ell = \bar{x}_\ell - \bar{x}$ of τ_ℓ always satisfy $\sum_{\ell=1}^{g} n_\ell \hat{\tau}_\ell = 0$. Under H_0, each $\hat{\tau}_\ell$ is an estimate of zero.) If the treatment contribution is large, H_0 should be rejected. The size of an array is quantified by stringing the rows of the array out into a vector and calculating its squared length. This quantity is called the *sum of squares* (SS). For the observations, we construct the vector $\mathbf{y}' = [9, 6, 9, 0, 2, 3, 1, 2]$. Its squared length is

$$SS_{\text{obs}} = 9^2 + 6^2 + 9^2 + 0^2 + 2^2 + 3^2 + 1^2 + 2^2 = 216$$

Similarly,

$$\begin{aligned} SS_{\text{mean}} &= 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 = 8(4^2) = 128 \\ SS_{\text{tr}} &= 4^2 + 4^2 + 4^2 + (-3)^2 + (-3)^2 + (-2)^2 + (-2)^2 + (-2)^2 \\ &= 3(4^2) + 2(-3)^2 + 3(-2)^2 = 78 \end{aligned}$$

and the residual sum of squares is

$$SS_{\text{res}} = 1^2 + (-2)^2 + 1^2 + (-1)^2 + 1^2 + 1^2 + (-1)^2 + 0^2 = 10$$

The sums of squares satisfy the same decomposition, (6-34), as the observations. Consequently,

$$SS_{\text{obs}} = SS_{\text{mean}} + SS_{\text{tr}} + SS_{\text{res}}$$

or $216 = 128 + 78 + 10$. The breakup into sums of squares apportions variability in the combined samples into mean, treatment, and residual (error) components. An analysis of variance proceeds by comparing the relative sizes of SS_{tr} and SS_{res}. If H_0 is true, variances computed from SS_{tr} and SS_{res} should be approximately equal. ■

The sum of squares decomposition illustrated numerically in Example 6.7 is so basic that the algebraic equivalent will now be developed.

Subtracting $\bar{x}$ from both sides of (6-34) and squaring gives

$$(x_{\ell j} - \bar{x})^2 = (\bar{x}_\ell - \bar{x})^2 + (x_{\ell j} - \bar{x}_\ell)^2 + 2(\bar{x}_\ell - \bar{x})(x_{\ell j} - \bar{x}_\ell)$$

We can sum both sides over j, note that $\sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell) = 0$, and obtain

$$\sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})^2 = n_\ell(\bar{x}_\ell - \bar{x})^2 + \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)^2$$

Next, summing both sides over ℓ we get

$$\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})^2 = \sum_{\ell=1}^{g} n_\ell(\bar{x}_\ell - \bar{x})^2 + \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)^2 \qquad (6\text{-}35)$$

$$\begin{pmatrix} SS_{cor} \\ \text{total (corrected) SS} \end{pmatrix} = \begin{pmatrix} SS_{tr} \\ \text{between (samples) SS} \end{pmatrix} + \begin{pmatrix} SS_{res} \\ \text{within (samples) SS} \end{pmatrix}$$

or

$$\begin{array}{ccccccc} \displaystyle\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} x_{\ell j}^2 & = & (n_1 + n_2 + \cdots + n_g)\bar{x}^2 & + & \displaystyle\sum_{\ell=1}^{g} n_\ell(\bar{x}_\ell - \bar{x})^2 & + & \displaystyle\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)^2 \\ (SS_{obs}) & = & (SS_{mean}) & + & (SS_{tr}) & + & (SS_{res}) \end{array} \qquad (6\text{-}36)$$

In the course of establishing (6-36), we have verified that the arrays representing the mean, treatment effects, and residuals are *orthogonal*. That is, these arrays, considered as vectors, are perpendicular whatever the observation vector $\mathbf{y}' = [x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{2n_2}, \ldots, x_{gn_g}]$. Consequently, we could obtain SS_{res} by subtraction, without having to calculate the individual residuals, because $SS_{res} = SS_{obs} - SS_{mean} - SS_{tr}$. However, this is false economy because plots of the residuals provide checks on the assumptions of the model.

The vector representations of the arrays involved in the decomposition (6-34) also have geometric interpretations that provide the degrees of freedom. For an arbitrary set of observations, let $[x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{2n_2}, \ldots, x_{gn_g}] = \mathbf{y}'$. The observation vector $\mathbf{y}$ can lie anywhere in $n = n_1 + n_2 + \cdots + n_g$ dimensions; the mean vector $\bar{x}\mathbf{1} = [\bar{x}, \ldots, \bar{x}]'$ must lie along the equiangular line of $\mathbf{1}$, and the treatment effect vector

$$(\bar{x}_1 - \bar{x}) \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_1 \\ \\ \\ \\ \\ \\ \\ \end{matrix} + (\bar{x}_2 - \bar{x}) \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_2 \\ \\ \\ \\ \end{matrix} + \cdots + (\bar{x}_g - \bar{x}) \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_g \end{matrix}$$

$$= (\bar{x}_1 - \bar{x})\mathbf{u}_1 + (\bar{x}_2 - \bar{x})\mathbf{u}_2 + \cdots + (\bar{x}_g - \bar{x})\mathbf{u}_g$$

lies in the hyperplane of linear combinations of the g vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_g$. Since $\mathbf{1} = \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_g$, the mean vector also lies in this hyperplane, and it is *always* perpendicular to the treatment vector. (See Exercise 6.10.) Thus, the mean vector has the freedom to lie anywhere along the one-dimensional equiangular line, and the treatment vector has the freedom to lie anywhere in the other $g - 1$ dimensions. The residual vector, $\hat{\mathbf{e}} = \mathbf{y} - (\bar{x}\mathbf{1}) - [(\bar{x}_1 - \bar{x})\mathbf{u}_1 + \cdots + (\bar{x}_g - \bar{x})\mathbf{u}_g]$ is perpendicular to both the mean vector and the treatment effect vector and has the freedom to lie anywhere in the subspace of dimension $n - (g - 1) - 1 = n - g$ that is perpendicular to their hyperplane.

To summarize, we attribute 1 d.f. to SS_{mean}, $g - 1$ d.f. to SS_{tr}, and $n - g = (n_1 + n_2 + \cdots + n_g) - g$ d.f. to SS_{res}. The total number of degrees of freedom is $n = n_1 + n_2 + \cdots + n_g$. Alternatively, by appealing to the univariate distribution theory, we find that these are the degrees of freedom for the chi-square distributions associated with the corresponding sums of squares.

The calculations of the sums of squares and the associated degrees of freedom are conveniently summarized by an ANOVA table.

ANOVA Table for Comparing Univariate Population Means

Source of variation	Sum of squares (SS)	Degrees of freedom (d.f.)
Treatments	$SS_{\text{tr}} = \sum_{\ell=1}^{g} n_\ell(\bar{x}_\ell - \bar{x})^2$	$g - 1$
Residual (error)	$SS_{\text{res}} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)^2$	$\sum_{\ell=1}^{g} n_\ell - g$
Total (corrected for the mean)	$SS_{\text{cor}} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})^2$	$\sum_{\ell=1}^{g} n_\ell - 1$

The usual F-test rejects H_0: $\tau_1 = \tau_2 = \cdots = \tau_g = 0$ at level α if

$$F = \frac{SS_{\text{tr}}/(g - 1)}{SS_{\text{res}}\Big/\left(\sum_{\ell=1}^{g} n_\ell - g\right)} > F_{g-1, \Sigma n_\ell - g}(\alpha)$$

where $F_{g-1, \Sigma n_\ell - g}(\alpha)$ is the upper (100α)th percentile of the F-distribution with $g - 1$ and $\Sigma n_\ell - g$ degrees of freedom. This is equivalent to rejecting H_0 for large values of $SS_{\text{tr}}/SS_{\text{res}}$ or for large values of $1 + SS_{\text{tr}}/SS_{\text{res}}$. The statistic appropriate for a multivariate generalization rejects H_0 for *small* values of the reciprocal

$$\frac{1}{1 + SS_{\text{tr}}/SS_{\text{res}}} = \frac{SS_{\text{res}}}{SS_{\text{res}} + SS_{\text{tr}}} \tag{6-37}$$

Example 6.8 (A univariate ANOVA table and F-test for treatment effects) Using the information in Example 6.7, we have the following ANOVA table:

Source of variation	Sum of squares	Degrees of freedom
Treatments	$SS_{tr} = 78$	$g - 1 = 3 - 1 = 2$
Residual	$SS_{res} = 10$	$\sum_{\ell=1}^{g} n_\ell - g = (3 + 2 + 3) - 3 = 5$
Total (corrected)	$SS_{cor} = 88$	$\sum_{\ell=1}^{g} n_\ell - 1 = 7$

Consequently,

$$F = \frac{SS_{tr}/(g-1)}{SS_{res}/(\Sigma n_\ell - g)} = \frac{78/2}{10/5} = 19.5$$

Since $F = 19.5 > F_{2,5}(.01) = 13.27$, we reject $H_0: \tau_1 = \tau_2 = \tau_3 = 0$ (no treatment effect) at the 1% level of significance. ■

Multivariate Analysis of Variance (MANOVA)

Paralleling the univariate reparameterization, we specify the MANOVA model:

MANOVA Model For Comparing g Population Mean Vectors

$$\mathbf{X}_{\ell j} = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell + \mathbf{e}_{\ell j}, \quad j = 1, 2, \ldots, n_\ell \quad \text{and} \quad \ell = 1, 2, \ldots, g \qquad (6\text{-}38)$$

where the $\mathbf{e}_{\ell j}$ are independent $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ variables. Here the parameter vector $\boldsymbol{\mu}$ is an overall mean (level), and $\boldsymbol{\tau}_\ell$ represents the ℓth treatment effect with $\sum_{\ell=1}^{g} n_\ell \boldsymbol{\tau}_\ell = \mathbf{0}$.

According to the model in (6-38), *each component* of the observation vector $\mathbf{X}_{\ell j}$ satisfies the univariate model (6-33). The errors for the components of $\mathbf{X}_{\ell j}$ are correlated, but the covariance matrix $\boldsymbol{\Sigma}$ is the same for all populations.

A vector of observations may be decomposed as suggested by the model. Thus,

$$\begin{array}{ccccccc} \mathbf{x}_{\ell j} & = & \bar{\mathbf{x}} & + & (\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}}) & + & (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell) \\ \text{(observation)} & & \begin{pmatrix}\text{overall sample}\\ \text{mean } \hat{\boldsymbol{\mu}}\end{pmatrix} & & \begin{pmatrix}\text{estimated}\\ \text{treatment}\\ \text{effect } \hat{\boldsymbol{\tau}}_\ell\end{pmatrix} & & \begin{pmatrix}\text{residual}\\ \hat{\mathbf{e}}_{\ell j}\end{pmatrix} \end{array} \qquad (6\text{-}39)$$

The decomposition in (6-39) leads to the multivariate analog of the univariate sum of squares breakup in (6-35). First we note that the product

$$(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})'$$

can be written as

$$
\begin{aligned}
(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})' &= [(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell) + (\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})][(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell) + (\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})]' \\
&= (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)' + (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})' \\
&\quad + (\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)' + (\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})'
\end{aligned}
$$

The sum over j of the middle two expressions is the zero matrix, because $\sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell) = \mathbf{0}$. Hence, summing the cross product over ℓ and j yields

$$
\underset{\begin{pmatrix}\text{total (corrected) sum} \\ \text{of squares and cross} \\ \text{products}\end{pmatrix}}{\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})'} = \underset{\begin{pmatrix}\text{treatment (Between)} \\ \text{sum of squares and} \\ \text{cross products}\end{pmatrix}}{\sum_{\ell=1}^{g} n_\ell(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})'} + \underset{\begin{pmatrix}\text{residual (Within) sum} \\ \text{of squares and cross} \\ \text{products}\end{pmatrix}}{\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)'} \tag{6-40}
$$

The *within* sum of squares and cross products matrix can be expressed as

$$
\begin{aligned}
\mathbf{W} &= \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)' \\
&= (n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + \cdots + (n_g - 1)\mathbf{S}_g
\end{aligned} \tag{6-41}
$$

where $\mathbf{S}_\ell$ is the sample covariance matrix for the ℓth sample. This matrix is a generalization of the $(n_1 + n_2 - 2)\mathbf{S}_{\text{pooled}}$ matrix encountered in the two-sample case. It plays a dominant role in testing for the presence of treatment effects.

Analogous to the univariate result, the hypothesis of no treatment effects,

$$H_0: \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = \mathbf{0}$$

is tested by considering the relative sizes of the treatment and residual sums of squares and cross products. Equivalently, we may consider the relative sizes of the residual and total (corrected) sum of squares and cross products. Formally, we summarize the calculations leading to the test statistic in a MANOVA table.

MANOVA Table for Comparing Population Mean Vectors

Source of variation	Matrix of sum of squares and cross products (SSP)	Degrees of freedom (d.f.)
Treatment	$\mathbf{B} = \sum_{\ell=1}^{g} n_\ell(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})'$	$g - 1$
Residual (Error)	$\mathbf{W} = \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)'$	$\sum_{\ell=1}^{g} n_\ell - g$
Total (corrected for the mean)	$\mathbf{B} + \mathbf{W} = \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})'$	$\sum_{\ell=1}^{g} n_\ell - 1$

This table is exactly the same form, component by component, as the ANOVA table, except that squares of scalars are replaced by their vector counterparts. For example, $(\bar{x}_\ell - \bar{x})^2$ becomes $(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})'$. The degrees of freedom correspond to the univariate geometry and also to some multivariate distribution theory involving Wishart densities. (See [1].)

One test of $H_0\colon \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = \mathbf{0}$ involves generalized variances. We reject H_0 if the ratio of generalized variances

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{\left| \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)' \right|}{\left| \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})' \right|} \tag{6-42}$$

is too small. The quantity $\Lambda^* = |\mathbf{W}|/|\mathbf{B} + \mathbf{W}|$, proposed originally by Wilks (see [25]), corresponds to the equivalent form (6-37) of the F-test of H_0: no treatment effects in the univariate case. *Wilks' lambda* has the virtue of being convenient and related to the likelihood ratio criterion.[2] The exact distribution of Λ^* can be derived for the special cases listed in Table 6.3. For other cases and large sample sizes, a modification of Λ^* due to Bartlett (see [4]) can be used to test H_0.

Table 6.3 Distribution of Wilks' Lambda, $\Lambda^* = |\mathbf{W}|/|\mathbf{B} + \mathbf{W}|$

No. of variables	No. of groups	Sampling distribution for multivariate normal data
$p = 1$	$g \ge 2$	$\left(\frac{\Sigma n_\ell - g}{g - 1}\right)\left(\frac{1 - \Lambda^*}{\Lambda^*}\right) \sim F_{g-1, \Sigma n_\ell - g}$
$p = 2$	$g \ge 2$	$\left(\frac{\Sigma n_\ell - g - 1}{g - 1}\right)\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1), 2(\Sigma n_\ell - g - 1)}$
$p \ge 1$	$g = 2$	$\left(\frac{\Sigma n_\ell - p - 1}{p}\right)\left(\frac{1 - \Lambda^*}{\Lambda^*}\right) \sim F_{p, \Sigma n_\ell - p - 1}$
$p \ge 1$	$g = 3$	$\left(\frac{\Sigma n_\ell - p - 2}{p}\right)\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p, 2(\Sigma n_\ell - p - 2)}$

[2] Wilks' lambda can also be expressed as a function of the eigenvalues of $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_s$ of $\mathbf{W}^{-1}\mathbf{B}$ as

$$\Lambda^* = \prod_{i=1}^{s} \left(\frac{1}{1 + \hat{\lambda}_i}\right)$$

where $s = \min(p, g - 1)$, the rank of $\mathbf{B}$. Other statistics for checking the equality of several multivariate means, such as Pillai's statistic, the Lawley–Hotelling statistic, and Roy's largest root statistic can also be written as particular functions of the eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$. For large samples, all of these statistics are, essentially equivalent. (See the additional discussion on page 336.)

Bartlett (see [4]) has shown that if H_0 is true and $\Sigma n_\ell = n$ is large,

$$-\left(n - 1 - \frac{(p+g)}{2}\right)\ln \Lambda^* = -\left(n - 1 - \frac{(p+g)}{2}\right)\ln\left(\frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|}\right) \tag{6-43}$$

has approximately a chi-square distribution with $p(g-1)$ d.f. Consequently, for $\Sigma n_\ell = n$ large, we reject H_0 at significance level α if

$$-\left(n - 1 - \frac{(p+g)}{2}\right)\ln\left(\frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|}\right) > \chi^2_{p(g-1)}(\alpha) \tag{6-44}$$

where $\chi^2_{p(g-1)}(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with $p(g-1)$ d.f.

Example 6.9 (A MANOVA table and Wilks' lambda for testing the equality of three mean vectors) Suppose an additional variable is observed along with the variable introduced in Example 6.7. The sample sizes are $n_1 = 3, n_2 = 2$, and $n_3 = 3$. Arranging the observation pairs $\mathbf{x}_{\ell j}$ in rows, we obtain

$$\begin{pmatrix} \begin{bmatrix} 9 \\ 3 \end{bmatrix} & \begin{bmatrix} 6 \\ 2 \end{bmatrix} & \begin{bmatrix} 9 \\ 7 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 4 \end{bmatrix} & \begin{bmatrix} 2 \\ 0 \end{bmatrix} & \\ \begin{bmatrix} 3 \\ 8 \end{bmatrix} & \begin{bmatrix} 1 \\ 9 \end{bmatrix} & \begin{bmatrix} 2 \\ 7 \end{bmatrix} \end{pmatrix} \qquad \begin{array}{l} \text{with } \bar{\mathbf{x}}_1 = \begin{bmatrix} 8 \\ 4 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \bar{\mathbf{x}}_3 = \begin{bmatrix} 2 \\ 8 \end{bmatrix}, \\ \text{and } \bar{\mathbf{x}} = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \end{array}$$

We have already expressed the observations on the first variable as the sum of an overall mean, treatment effect, and residual in our discussion of univariate ANOVA. We found that

$$\underset{\text{(observation)}}{\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & \\ 3 & 1 & 2 \end{pmatrix}} = \underset{\text{(mean)}}{\begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 & \\ 4 & 4 & 4 \end{pmatrix}} + \underset{\begin{pmatrix}\text{treatment}\\ \text{effect}\end{pmatrix}}{\begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 & \\ -2 & -2 & -2 \end{pmatrix}} + \underset{\text{(residual)}}{\begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & \\ 1 & -1 & 0 \end{pmatrix}}$$

and

$$\begin{aligned} SS_{obs} &= SS_{mean} + SS_{tr} + SS_{res} \\ 216 &= 128 + 78 + 10 \\ \text{Total SS (corrected)} &= SS_{obs} - SS_{mean} = 216 - 128 = 88 \end{aligned}$$

Repeating this operation for the observations on the second variable, we have

$$\underset{\text{(observation)}}{\begin{pmatrix} 3 & 2 & 7 \\ 4 & 0 & \\ 8 & 9 & 7 \end{pmatrix}} = \underset{\text{(mean)}}{\begin{pmatrix} 5 & 5 & 5 \\ 5 & 5 & \\ 5 & 5 & 5 \end{pmatrix}} + \underset{\begin{pmatrix}\text{treatment}\\ \text{effect}\end{pmatrix}}{\begin{pmatrix} -1 & -1 & -1 \\ -3 & -3 & \\ 3 & 3 & 3 \end{pmatrix}} + \underset{\text{(residual)}}{\begin{pmatrix} -1 & -2 & 3 \\ 2 & -2 & \\ 0 & 1 & -1 \end{pmatrix}}$$

and

$$\begin{aligned} SS_{\text{obs}} &= SS_{\text{mean}} + SS_{\text{tr}} + SS_{\text{res}} \\ 272 &= 200 + 48 + 24 \\ \text{Total SS (corrected)} &= SS_{\text{obs}} - SS_{\text{mean}} = 272 - 200 = 72 \end{aligned}$$

These two single-component analyses must be augmented with the sum of entry-by-entry *cross products* in order to complete the entries in the MANOVA table. Proceeding row by row in the arrays for the two variables, we obtain the cross product contributions:

$$\begin{aligned} \text{Mean:}\ & 4(5) + 4(5) + \cdots + 4(5) = 8(4)(5) = 160 \\ \text{Treatment:}\ & 3(4)(-1) + 2(-3)(-3) + 3(-2)(3) = -12 \\ \text{Residual:}\ & 1(-1) + (-2)(-2) + 1(3) + (-1)(2) + \cdots + 0(-1) = 1 \\ \text{Total:}\ & 9(3) + 6(2) + 9(7) + 0(4) + \cdots + 2(7) = 149 \end{aligned}$$

$$\begin{aligned} \text{Total (corrected) cross product} &= \text{total cross product} - \text{mean cross product} \\ &= 149 - 160 = -11 \end{aligned}$$

Thus, the MANOVA table takes the following form:

Source of variation	Matrix of sum of squares and cross products	Degrees of freedom
Treatment	$\begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix}$	$3 - 1 = 2$
Residual	$\begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$	$3 + 2 + 3 - 3 = 5$
Total (corrected)	$\begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix}$	7

Equation (6-40) is verified by noting that

$$\begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix} = \begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix} + \begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$$

Using (6-42), we get

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{\begin{vmatrix} 10 & 1 \\ 1 & 24 \end{vmatrix}}{\begin{vmatrix} 88 & -11 \\ -11 & 72 \end{vmatrix}} = \frac{10(24) - (1)^2}{88(72) - (-11)^2} = \frac{239}{6215} = .0385$$

Since $p = 2$ and $g = 3$, Table 6.3 indicates that an exact test (assuming normality and equal group covariance matrices) of H_0: $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \boldsymbol{\tau}_3 = \mathbf{0}$ (no treatment effects) versus H_1: at least one $\boldsymbol{\tau}_\ell \neq \mathbf{0}$ is available. To carry out the test, we compare the test statistic

$$\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right)\frac{(\Sigma n_\ell - g - 1)}{(g - 1)} = \left(\frac{1 - \sqrt{.0385}}{\sqrt{.0385}}\right)\left(\frac{8 - 3 - 1}{3 - 1}\right) = 8.19$$

with a percentage point of an F-distribution having $\nu_1 = 2(g - 1) = 4$ and $\nu_2 = 2(\Sigma n_\ell - g - 1) = 8$ d.f. Since $8.19 > F_{4,8}(.01) = 7.01$, we reject H_0 at the $\alpha = .01$ level and conclude that treatment differences exist. ■

When the number of variables, p, is large, the MANOVA table is usually not constructed. Still, it is good practice to have the computer print the matrices $\mathbf{B}$ and $\mathbf{W}$ so that especially large entries can be located. Also, the residual vectors

$$\hat{\mathbf{e}}_{\ell j} = \mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell$$

should be examined for normality and the presence of outliers using the techniques discussed in Sections 4.6 and 4.7 of Chapter 4.

Example 6.10 (A multivariate analysis of Wisconsin nursing home data) The Wisconsin Department of Health and Social Services reimburses nursing homes in the state for the services provided. The department develops a set of formulas for rates for each facility, based on factors such as level of care, mean wage rate, and average wage rate in the state.

Nursing homes can be classified on the basis of ownership (private party, nonprofit organization, and government) and certification (skilled nursing facility, intermediate care facility, or a combination of the two).

One purpose of a recent study was to investigate the effects of ownership or certification (or both) on costs. Four costs, computed on a per-patient-day basis and measured in hours per patient day, were selected for analysis: X_1 = cost of nursing labor, X_2 = cost of dietary labor, X_3 = cost of plant operation and maintenance labor, and X_4 = cost of housekeeping and laundry labor. A total of $n = 516$ observations on each of the $p = 4$ cost variables were initially separated according to ownership. Summary statistics for each of the $g = 3$ groups are given in the following table.

Group	Number of observations	Sample mean vectors
$\ell = 1$ (private)	$n_1 = 271$	
$\ell = 2$ (nonprofit)	$n_2 = 138$	$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2.066 \\ .480 \\ .082 \\ .360 \end{bmatrix}$; $\bar{\mathbf{x}}_2 = \begin{bmatrix} 2.167 \\ .596 \\ .124 \\ .418 \end{bmatrix}$; $\bar{\mathbf{x}}_3 = \begin{bmatrix} 2.273 \\ .521 \\ .125 \\ .383 \end{bmatrix}$
$\ell = 3$ (government)	$n_3 = 107$	
	$\sum_{\ell=1}^{3} n_\ell = 516$	

Sample covariance matrices

$$\mathbf{S}_1 = \begin{bmatrix} .291 & & & \\ -.001 & .011 & & \\ .002 & .000 & .001 & \\ .010 & .003 & .000 & .010 \end{bmatrix}; \qquad \mathbf{S}_2 = \begin{bmatrix} .561 & & & \\ .011 & .025 & & \\ .001 & .004 & .005 & \\ .037 & .007 & .002 & .019 \end{bmatrix};$$

$$\mathbf{S}_3 = \begin{bmatrix} .261 & & & \\ .030 & .017 & & \\ .003 & -.000 & .004 & \\ .018 & .006 & .001 & .013 \end{bmatrix}$$

Source: Data courtesy of State of Wisconsin Department of Health and Social Services.

Since the $\mathbf{S}_\ell$'s seem to be reasonably compatible,[3] they were pooled [see (6-41)] to obtain

$$\mathbf{W} = (n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + (n_3 - 1)\mathbf{S}_3$$

$$= \begin{bmatrix} 182.962 & & & \\ 4.408 & 8.200 & & \\ 1.695 & .633 & 1.484 & \\ 9.581 & 2.428 & .394 & 6.538 \end{bmatrix}$$

Also,

$$\bar{\mathbf{x}} = \frac{n_1\bar{\mathbf{x}}_1 + n_2\bar{\mathbf{x}}_2 + n_3\bar{\mathbf{x}}_3}{n_1 + n_2 + n_3} = \begin{bmatrix} 2.136 \\ .519 \\ .102 \\ .380 \end{bmatrix}$$

and

$$\mathbf{B} = \sum_{\ell=1}^{3} n_\ell(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})' = \begin{bmatrix} 3.475 & & & \\ 1.111 & 1.225 & & \\ .821 & .453 & .235 & \\ .584 & .610 & .230 & .304 \end{bmatrix}$$

To test $H_0\colon \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \boldsymbol{\tau}_3$ (no ownership effects or, equivalently, no difference in average costs among the three types of owners—private, nonprofit, and government), we can use the result in Table 6.3 for $g = 3$.

Computer-based calculations give

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = .7714$$

[3]However, a normal-theory test of $H_0\colon \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3$ would reject H_0 at any reasonable significance level because of the large sample sizes (see Example 6.12).

and

$$\left(\frac{\Sigma n_\ell - p - 2}{p}\right)\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) = \left(\frac{516 - 4 - 2}{4}\right)\left(\frac{1 - \sqrt{.7714}}{\sqrt{.7714}}\right) = 17.67$$

Let $\alpha = .01$, so that $F_{2(4),2(510)}(.01) \doteq \chi_8^2(.01)/8 = 2.51$. Since $17.67 > F_{8,1020}(.01) \doteq 2.51$, we reject H_0 at the 1% level and conclude that average costs differ, depending on type of ownership.

It is informative to compare the results based on this "exact" test with those obtained using the large-sample procedure summarized in (6-43) and (6-44). For the present example, $\Sigma n_\ell = n = 516$ is large, and H_0 can be tested at the $\alpha = .01$ level by comparing

$$-(n - 1 - (p + g)/2) \ln\left(\frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|}\right) = -511.5 \ln(.7714) = 132.76$$

with $\chi^2_{p(g-1)}(.01) = \chi_8^2(.01) = 20.09$. Since $132.76 > \chi_8^2(.01) = 20.09$, we reject H_0 at the 1% level. This result is consistent with the result based on the foregoing F-statistic. ■

6.5 Simultaneous Confidence Intervals for Treatment Effects

When the hypothesis of equal treatment effects is rejected, those effects that led to the rejection of the hypothesis are of interest. For pairwise comparisons, the Bonferroni approach (see Section 5.4) can be used to construct simultaneous confidence intervals for the components of the differences $\boldsymbol{\tau}_k - \boldsymbol{\tau}_\ell$ (or $\boldsymbol{\mu}_k - \boldsymbol{\mu}_\ell$). These intervals are shorter than those obtained for all contrasts, and they require critical values only for the univariate t-statistic.

Let τ_{ki} be the ith component of $\boldsymbol{\tau}_k$. Since $\boldsymbol{\tau}_k$ is estimated by $\hat{\boldsymbol{\tau}}_k = \bar{\mathbf{x}}_k - \bar{\mathbf{x}}$

$$\hat{\tau}_{ki} = \bar{x}_{ki} - \bar{x}_i \tag{6-45}$$

and $\hat{\tau}_{ki} - \hat{\tau}_{\ell i} = \bar{x}_{ki} - \bar{x}_{\ell i}$ is the difference between two independent sample means. The two-sample t-based confidence interval is valid with an appropriately modified α. Notice that

$$\text{Var}(\hat{\tau}_{ki} - \hat{\tau}_{\ell i}) = \text{Var}(\bar{X}_{ki} - \bar{X}_{\ell i}) = \left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)\sigma_{ii}$$

where σ_{ii} is the ith diagonal element of $\boldsymbol{\Sigma}$. As suggested by (6-41), $\text{Var}(\bar{X}_{ki} - \bar{X}_{\ell i})$ is estimated by dividing the corresponding element of $\mathbf{W}$ by its degrees of freedom. That is,

$$\widehat{\text{Var}}(\bar{X}_{ki} - \bar{X}_{\ell i}) = \left(\frac{1}{n_k} + \frac{1}{n_\ell}\right)\frac{w_{ii}}{n - g}$$

where w_{ii} is the ith diagonal element of $\mathbf{W}$ and $n = n_1 + \cdots + n_g$.

It remains to apportion the error rate over the numerous confidence statements. Relation (5-28) still applies. There are p variables and $g(g-1)/2$ pairwise differences, so each two-sample t-interval will employ the critical value $t_{n-g}(\alpha/2m)$, where

$$m = pg(g-1)/2 \tag{6-46}$$

is the number of simultaneous confidence statements.

Result 6.5. Let $n = \sum_{k=1}^{g} n_k$. For the model in (6-38), with confidence at least $(1-\alpha)$,

$$\tau_{ki} - \tau_{\ell i} \quad \text{belongs to} \quad \bar{x}_{ki} - \bar{x}_{\ell i} \pm t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_k}+\frac{1}{n_\ell}\right)}$$

for all components $i = 1, \ldots, p$ and all differences $\ell < k = 1, \ldots, g$. Here w_{ii} is the ith diagonal element of $\mathbf{W}$.

We shall illustrate the construction of simultaneous interval estimates for the pairwise differences in treatment means using the nursing-home data introduced in Example 6.10.

Example 6.11 (Simultaneous intervals for treatment differences—nursing homes) We saw in Example 6.10 that average costs for nursing homes differ, depending on the type of ownership. We can use Result 6.5 to estimate the magnitudes of the differences. A comparison of the variable X_3, costs of plant operation and maintenance labor, between privately owned nursing homes and government-owned nursing homes can be made by estimating $\tau_{13} - \tau_{33}$. Using (6-39) and the information in Example 6.10, we have

$$\hat{\boldsymbol{\tau}}_1 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}) = \begin{bmatrix} -.070 \\ -.039 \\ -.020 \\ -.020 \end{bmatrix}, \qquad \hat{\boldsymbol{\tau}}_3 = (\bar{\mathbf{x}}_3 - \bar{\mathbf{x}}) = \begin{bmatrix} .137 \\ .002 \\ .023 \\ .003 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 182.962 & & & \\ 4.408 & 8.200 & & \\ 1.695 & .633 & 1.484 & \\ 9.581 & 2.428 & .394 & 6.538 \end{bmatrix}$$

Consequently,

$$\hat{\tau}_{13} - \hat{\tau}_{33} = -.020 - .023 = -.043$$

and $n = 271 + 138 + 107 = 516$, so that

$$\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_3}\right)\frac{w_{33}}{n-g}} = \sqrt{\left(\frac{1}{271}+\frac{1}{107}\right)\frac{1.484}{516-3}} = .00614$$

Since $p = 4$ and $g = 3$, for 95% simultaneous confidence statements we require that $t_{513}(.05/4(3)2) \doteq 2.87$. (See Appendix, Table 1.) The 95% simultaneous confidence statement is

$$\begin{aligned}\tau_{13} - \tau_{33} \text{ belongs to } \hat{\tau}_{13} - \hat{\tau}_{33} &\pm t_{513}(.00208)\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_3}\right)\frac{w_{33}}{n-g}} \\ &= -.043 \pm 2.87(.00614) \\ &= -.043 \pm .018, \text{ or } (-.061, -.025)\end{aligned}$$

We conclude that the average maintenance and labor cost for government-owned nursing homes is higher by .025 to .061 hour per patient day than for privately owned nursing homes. With the same 95% confidence, we can say that

$$\tau_{13} - \tau_{23} \text{ belongs to the interval } (-.058, -.026)$$

and

$$\tau_{23} - \tau_{33} \text{ belongs to the interval } (-.021, .019)$$

Thus, a difference in this cost exists between private and nonprofit nursing homes, but no difference is observed between nonprofit and government nursing homes. ■

6.6 Testing for Equality of Covariance Matrices

One of the assumptions made when comparing two or more multivariate mean vectors is that the covariance matrices of the potentially different populations are the same. (This assumption will appear again in Chapter 11 when we discuss discrimination and classification.) Before pooling the variation across samples to form a pooled covariance matrix when comparing mean vectors, it can be worthwhile to test the equality of the population covariance matrices. One commonly employed test for equal covariance matrices is Box's M-test ([8], [9]).

With g populations, the null hypothesis is

$$H_0: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma} \tag{6-47}$$

where $\boldsymbol{\Sigma}_\ell$ is the covariance matrix for the ℓth population, $\ell = 1, 2, \ldots, g$, and $\boldsymbol{\Sigma}$ is the presumed common covariance matrix. The alternative hypothesis is that at least two of the covariance matrices are not equal.

Assuming multivariate normal populations, a likelihood ratio statistic for testing (6-47) is given by (see [1])

$$\Lambda = \prod_\ell \left(\frac{|\mathbf{S}_\ell|}{|\mathbf{S}_{\text{pooled}}|}\right)^{(n_\ell - 1)/2} \tag{6-48}$$

Here n_ℓ is the sample size for the ℓth group, $\mathbf{S}_\ell$ is the ℓth group sample covariance matrix and $\mathbf{S}_{\text{pooled}}$ is the pooled sample covariance matrix given by

$$\mathbf{S}_{\text{pooled}} = \frac{1}{\sum_\ell (n_\ell - 1)}\left\{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + \cdots + (n_g - 1)\mathbf{S}_g\right\} \tag{6-49}$$

Box's test is based on his χ^2 approximation to the sampling distribution of $-2 \ln \Lambda$ (see Result 5.2). Setting $-2 \ln \Lambda = M$ (Box's M statistic) gives

$$M = \left[\sum_{\ell} (n_\ell - 1) \right] \ln |\mathbf{S}_{\text{pooled}}| - \sum_{\ell} [(n_\ell - 1) \ln |\mathbf{S}_\ell|] \tag{6-50}$$

If the null hypothesis is true, the individual sample covariance matrices are not expected to differ too much and, consequently, do not differ too much from the pooled covariance matrix. In this case, the ratio of the determinants in (6-48) will all be close to 1, Λ will be near 1 and Box's M statistic will be small. If the null hypothesis is false, the sample covariance matrices can differ more and the differences in their determinants will be more pronounced. In this case Λ will be small and M will be relatively large. To illustrate, note that the determinant of the pooled covariance matrix, $|\mathbf{S}_{\text{pooled}}|$, will lie somewhere near the "middle" of the determinants $|\mathbf{S}_\ell|$'s of the individual group covariance matrices. As the latter quantities become more disparate, the product of the ratios in (6-44) will get closer to 0. In fact, as the $|\mathbf{S}_\ell|$'s increase in spread, $|\mathbf{S}_{(1)}|/|\mathbf{S}_{\text{pooled}}|$ reduces the product proportionally more than $|\mathbf{S}_{(g)}|/|\mathbf{S}_{\text{pooled}}|$ increases it, where $|\mathbf{S}_{(1)}|$ and $|\mathbf{S}_{(g)}|$ are the minimum and maximum determinant values, respectively.

Box's Test for Equality of Covariance Matrices

Set

$$u = \left[\sum_{\ell} \frac{1}{(n_\ell - 1)} - \frac{1}{\sum_{\ell} (n_\ell - 1)} \right] \left[\frac{2p^2 + 3p - 1}{6(p + 1)(g - 1)} \right] \tag{6-51}$$

where p is the number of variables and g is the number of groups. Then

$$C = (1 - u)M = (1 - u)\left\{ \left[\sum_{\ell} (n_\ell - 1) \right] \ln |\mathbf{S}_{\text{pooled}}| - \sum_{\ell} [(n_\ell - 1) \ln |\mathbf{S}_\ell|] \right\} \tag{6-52}$$

has an approximate χ^2 distribution with

$$v = g\frac{1}{2}p(p + 1) - \frac{1}{2}p(p + 1) = \frac{1}{2}p(p + 1)(g - 1) \tag{6-53}$$

degrees of freedom. At significance level α, reject H_0 if $C > \chi^2_{p(p+1)(g-1)/2}(\alpha)$.

Box's χ^2 approximation works well if each n_ℓ exceeds 20 and if p and g do not exceed 5. In situations where these conditions do not hold, Box ([7], [8]) has provided a more precise F approximation to the sampling distribution of M.

Example 6.12 (Testing equality of covariance matrices—nursing homes) We introduced the Wisconsin nursing home data in Example 6.10. In that example the sample covariance matrices for $p = 4$ cost variables associated with $g = 3$ groups of nursing homes are displayed. Assuming multivariate normal data, we test the hypothesis $H_0: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}$.

Using the information in Example 6.10, we have $n_1 = 271$, $n_2 = 138$, $n_3 = 107$ and $|\mathbf{S}_1| = 2.783 \times 10^{-8}$, $|\mathbf{S}_2| = 89.539 \times 10^{-8}$, $|\mathbf{S}_3| = 14.579 \times 10^{-8}$, and $|\mathbf{S}_{\text{pooled}}| = 17.398 \times 10^{-8}$. Taking the natural logarithms of the determinants gives $\ln|\mathbf{S}_1| = -17.397$, $\ln|\mathbf{S}_2| = -13.926$, $\ln|\mathbf{S}_3| = -15.741$ and $\ln|\mathbf{S}_{\text{pooled}}| = -15.564$. We calculate

$$u = \left[\frac{1}{270} + \frac{1}{137} + \frac{1}{106} - \frac{1}{270 + 137 + 106}\right]\left[\frac{2(4^2) + 3(4) - 1}{6(4 + 1)(3 - 1)}\right] = .0133$$

$$M = [270 + 137 + 106](-15.564) - [270(-17.397) + 137(-13.926) + 106(-15.741)]$$
$$= 289.3$$

and $C = (1 - .0133)289.3 = 285.5$. Referring C to a χ^2 table with $\nu = 4(4 + 1)(3 - 1)/2 = 20$ degrees of freedom, it is clear that H_0 is rejected at any reasonable level of significance. We conclude that the covariance matrices of the cost variables associated with the three populations of nursing homes are not the same. ■

Box's M-test is routinely calculated in many statistical computer packages that do MANOVA and other procedures requiring equal covariance matrices. It is known that the M-test is sensitive to some forms of non-normality. More broadly, in the presence of non-normality, normal theory tests on covariances are influenced by the kurtosis of the parent populations (see [16]). However, with reasonably large samples, the MANOVA tests of means or treatment effects are rather robust to nonnormality. Thus the M-test may reject H_0 in some non-normal cases where it is not damaging to the MANOVA tests. Moreover, with equal sample sizes, some differences in covariance matrices have little effect on the MANOVA tests. To summarize, we may decide to continue with the usual MANOVA tests even though the M-test leads to rejection of H_0.

6.7 Two-Way Multivariate Analysis of Variance

Following our approach to the one-way MANOVA, we shall briefly review the analysis for a *univariate* two-way fixed-effects model and then simply generalize to the multivariate case by analogy.

Univariate Two-Way Fixed-Effects Model with Interaction

We assume that measurements are recorded at various levels of two factors. In some cases, these experimental conditions represent levels of a single treatment arranged within several blocks. The particular experimental design employed will not concern us in this book. (See [10] and [17] for discussions of experimental design.) We shall, however, assume that observations at different combinations of experimental conditions are independent of one another.

Let the two sets of experimental conditions be the levels of, for instance, factor 1 and factor 2, respectively.[4] Suppose there are g levels of factor 1 and b levels of factor 2, and that n independent observations can be observed at each of the gb combi-

[4] The use of the term "factor" to indicate an experimental condition is convenient. The factors discussed here should not be confused with the unobservable factors considered in Chapter 9 in the context of factor analysis.

nations of levels. Denoting the rth observation at level ℓ of factor 1 and level k of factor 2 by $X_{\ell kr}$, we specify the univariate two-way model as

$$\begin{aligned} X_{\ell kr} &= \mu + \tau_\ell + \beta_k + \gamma_{\ell k} + e_{\ell kr} \\ \ell &= 1, 2, \ldots, g \\ k &= 1, 2, \ldots, b \\ r &= 1, 2, \ldots, n \end{aligned} \tag{6-54}$$

where $\sum_{\ell=1}^{g} \tau_\ell = \sum_{k=1}^{b} \beta_k = \sum_{\ell=1}^{g} \gamma_{\ell k} = \sum_{k=1}^{b} \gamma_{\ell k} = 0$ and the $e_{\ell kr}$ are independent $N(0, \sigma^2)$ random variables. Here μ represents an overall level, τ_ℓ represents the fixed effect of factor 1, β_k represents the fixed effect of factor 2, and $\gamma_{\ell k}$ is the interaction between factor 1 and factor 2. The expected response at the ℓth level of factor 1 and the kth level of factor 2 is thus

$$\begin{array}{ccccccccc} E(X_{\ell kr}) & = & \mu & + & \tau_\ell & + & \beta_k & + & \gamma_{\ell k} \\ \begin{pmatrix} \text{mean} \\ \text{response} \end{pmatrix} & = & \begin{pmatrix} \text{overall} \\ \text{level} \end{pmatrix} & + & \begin{pmatrix} \text{effect of} \\ \text{factor 1} \end{pmatrix} & + & \begin{pmatrix} \text{effect of} \\ \text{factor 2} \end{pmatrix} & + & \begin{pmatrix} \text{factor 1-factor 2} \\ \text{interaction} \end{pmatrix} \end{array}$$

$$\ell = 1, 2, \ldots, g, \qquad k = 1, 2, \ldots, b \tag{6-55}$$

The presence of interaction, $\gamma_{\ell k}$, implies that the factor effects are not additive and complicates the interpretation of the results. Figures 6.3(a) and (b) show

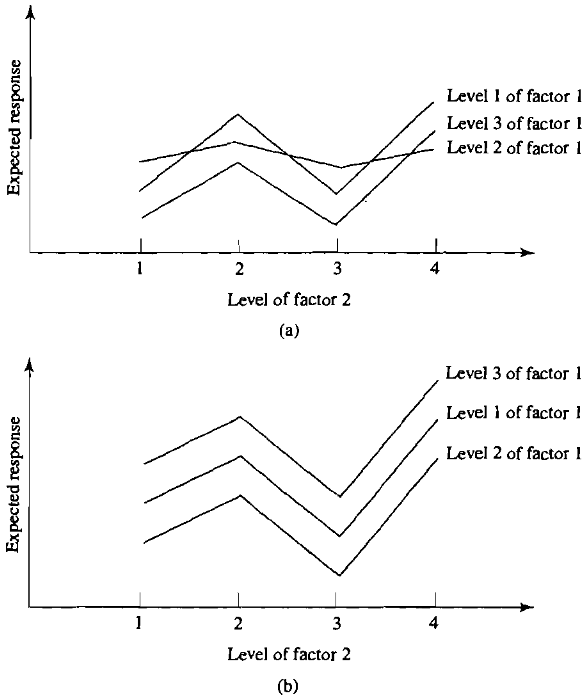


Figure 6.3 Curves for expected responses (a) with interaction and (b) without interaction.

expected responses as a function of the factor levels with and without interaction, respectively. The absense of interaction means $\gamma_{\ell k} = 0$ for all ℓ and k.

In a manner analogous to (6-55), each observation can be decomposed as

$$x_{\ell k r} = \bar{x} + (\bar{x}_{\ell\cdot} - \bar{x}) + (\bar{x}_{\cdot k} - \bar{x}) + (\bar{x}_{\ell k} - \bar{x}_{\ell\cdot} - \bar{x}_{\cdot k} + \bar{x}) + (x_{\ell k r} - \bar{x}_{\ell k}) \qquad (6\text{-}56)$$

where $\bar{x}$ is the overall average, $\bar{x}_{\ell\cdot}$ is the average for the ℓth level of factor 1, $\bar{x}_{\cdot k}$ is the average for the kth level of factor 2, and $\bar{x}_{\ell k}$ is the average for the ℓth level of factor 1 *and* the kth level of factor 2. Squaring and summing the deviations $(x_{\ell k r} - \bar{x})$ gives

$$\begin{aligned}\sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(x_{\ell k r} - \bar{x})^2 &= \sum_{\ell=1}^{g} bn(\bar{x}_{\ell\cdot} - \bar{x})^2 + \sum_{k=1}^{b} gn(\bar{x}_{\cdot k} - \bar{x})^2 \\ &\quad + \sum_{\ell=1}^{g}\sum_{k=1}^{b} n(\bar{x}_{\ell k} - \bar{x}_{\ell\cdot} - \bar{x}_{\cdot k} + \bar{x})^2 \\ &\quad + \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(x_{\ell k r} - \bar{x}_{\ell k})^2 \end{aligned} \qquad (6\text{-}57)$$

or

$$SS_{\text{cor}} = SS_{\text{fac1}} + SS_{\text{fac2}} + SS_{\text{int}} + SS_{\text{res}}$$

The corresponding degrees of freedom associated with the sums of squares in the breakup in (6-57) are

$$gbn - 1 = (g - 1) + (b - 1) + (g - 1)(b - 1) + gb(n - 1) \qquad (6\text{-}58)$$

The ANOVA table takes the following form:

ANOVA Table for Comparing Effects of Two Factors and Their Interaction

Source of variation	Sum of squares (SS)	Degrees of freedom (d.f.)
Factor 1	$SS_{\text{fac1}} = \sum_{\ell=1}^{g} bn(\bar{x}_{\ell\cdot} - \bar{x})^2$	$g - 1$
Factor 2	$SS_{\text{fac2}} = \sum_{k=1}^{b} gn(\bar{x}_{\cdot k} - \bar{x})^2$	$b - 1$
Interaction	$SS_{\text{int}} = \sum_{\ell=1}^{g}\sum_{k=1}^{b} n(\bar{x}_{\ell k} - \bar{x}_{\ell\cdot} - \bar{x}_{\cdot k} + \bar{x})^2$	$(g - 1)(b - 1)$
Residual (Error)	$SS_{\text{res}} = \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(x_{\ell k r} - \bar{x}_{\ell k})^2$	$gb(n - 1)$
Total (corrected)	$SS_{\text{cor}} = \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(x_{\ell k r} - \bar{x})^2$	$gbn - 1$

The F-ratios of the mean squares, $SS_{fac1}/(g-1)$, $SS_{fac2}/(b-1)$, and $SS_{int}/(g-1)(b-1)$ to the mean square, $SS_{res}/(gb(n-1))$ can be used to test for the effects of factor 1, factor 2, and factor 1–factor 2 interaction, respectively. (See [11] for a discussion of univariate two-way analysis of variance.)

Multivariate Two-Way Fixed-Effects Model with Interaction

Proceeding by analogy, we specify the two-way fixed-effects model for a *vector* response consisting of p components [see (6-54)]

$$\begin{aligned} \mathbf{X}_{\ell kr} &= \boldsymbol{\mu} + \boldsymbol{\tau}_\ell + \boldsymbol{\beta}_k + \boldsymbol{\gamma}_{\ell k} + \mathbf{e}_{\ell kr} \\ \ell &= 1, 2, \ldots, g \\ k &= 1, 2, \ldots, b \\ r &= 1, 2, \ldots, n \end{aligned} \tag{6-59}$$

where $\sum_{\ell=1}^{g} \boldsymbol{\tau}_\ell = \sum_{k=1}^{b} \boldsymbol{\beta}_k = \sum_{\ell=1}^{g} \boldsymbol{\gamma}_{\ell k} = \sum_{k=1}^{b} \boldsymbol{\gamma}_{\ell k} = \mathbf{0}$. The vectors are all of order $p \times 1$, and the $\mathbf{e}_{\ell kr}$ are independent $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ random vectors. Thus, the responses consist of p measurements replicated n times at each of the possible combinations of levels of factors 1 and 2.

Following (6-56), we can decompose the observation vectors $\mathbf{x}_{\ell kr}$ as

$$\mathbf{x}_{\ell kr} = \bar{\mathbf{x}} + (\bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}}) + (\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}}) + (\bar{\mathbf{x}}_{\ell k} - \bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}}) + (\mathbf{x}_{\ell kr} - \bar{\mathbf{x}}_{\ell k}) \tag{6-60}$$

where $\bar{\mathbf{x}}$ is the overall average of the observation vectors, $\bar{\mathbf{x}}_{\ell\cdot}$ is the average of the observation vectors at the ℓth level of factor 1, $\bar{\mathbf{x}}_{\cdot k}$ is the average of the observation vectors at the kth level of factor 2, and $\bar{\mathbf{x}}_{\ell k}$ is the average of the observation vectors at the ℓth level of factor 1 *and* the kth level of factor 2.

Straightforward generalizations of (6-57) and (6-58) give the breakups of the sum of squares and cross products and degrees of freedom:

$$\begin{aligned} \sum_{\ell=1}^{g} \sum_{k=1}^{b} \sum_{r=1}^{n} (\mathbf{x}_{\ell kr} - \bar{\mathbf{x}})(\mathbf{x}_{\ell kr} - \bar{\mathbf{x}})' &= \sum_{\ell=1}^{g} bn(\bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}})' \\ &\quad + \sum_{k=1}^{b} gn(\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}})' \\ &\quad + \sum_{\ell=1}^{g} \sum_{k=1}^{b} n(\bar{\mathbf{x}}_{\ell k} - \bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\ell k} - \bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}})' \\ &\quad + \sum_{\ell=1}^{g} \sum_{k=1}^{b} \sum_{r=1}^{n} (\mathbf{x}_{\ell kr} - \bar{\mathbf{x}}_{\ell k})(\mathbf{x}_{\ell kr} - \bar{\mathbf{x}}_{\ell k})' \end{aligned} \tag{6-61}$$

$$gbn - 1 = (g-1) + (b-1) + (g-1)(b-1) + gb(n-1) \tag{6-62}$$

Again, the generalization from the univariate to the multivariate analysis consists simply of replacing a scalar such as $(\bar{x}_{\ell\cdot} - \bar{x})^2$ with the corresponding matrix $(\bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}})'$.

The MANOVA table is the following:

MANOVA Table for Comparing Factors and Their Interaction

Source of variation	Matrix of sum of squares and cross products (SSP)	Degrees of freedom (d.f.)
Factor 1	$\mathbf{SSP}_{\text{fac}1} = \sum_{\ell=1}^{g} bn(\bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}})'$	$g - 1$
Factor 2	$\mathbf{SSP}_{\text{fac}2} = \sum_{k=1}^{b} gn(\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}})'$	$b - 1$
Interaction	$\mathbf{SSP}_{\text{int}} = \sum_{\ell=1}^{g}\sum_{k=1}^{b} n(\bar{\mathbf{x}}_{\ell k} - \bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\ell k} - \bar{\mathbf{x}}_{\ell\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}})'$	$(g - 1)(b - 1)$
Residual (Error)	$\mathbf{SSP}_{res} = \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n} (\mathbf{x}_{\ell kr} - \bar{\mathbf{x}}_{\ell k})(\mathbf{x}_{\ell kr} - \bar{\mathbf{x}}_{\ell k})'$	$gb(n - 1)$
Total (corrected)	$\mathbf{SSP}_{\text{cor}} = \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n} (\mathbf{x}_{\ell kr} - \bar{\mathbf{x}})(\mathbf{x}_{\ell kr} - \bar{\mathbf{x}})'$	$gbn - 1$

A test (the likelihood ratio test)[5] of

$$H_0: \boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = \cdots = \boldsymbol{\gamma}_{gb} = \mathbf{0} \qquad \text{(no interaction effects)} \tag{6-63}$$

versus

$$H_1: \text{At least one } \boldsymbol{\gamma}_{\ell k} \neq \mathbf{0}$$

is conducted by rejecting H_0 for small values of the ratio

$$\Lambda^* = \frac{|\mathbf{SSP}_{\text{res}}|}{|\mathbf{SSP}_{\text{int}} + \mathbf{SSP}_{\text{res}}|} \tag{6-64}$$

For large samples, Wilks' lambda, Λ^*, can be referred to a chi-square percentile. Using Bartlett's multiplier (see [6]) to improve the chi-square approximation, we reject $H_0: \boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = \cdots = \boldsymbol{\gamma}_{gb} = \mathbf{0}$ at the α level if

$$-\left[gb(n - 1) - \frac{p + 1 - (g - 1)(b - 1)}{2}\right] \ln \Lambda^* > \chi^2_{(g-1)(b-1)p}(\alpha) \tag{6-65}$$

where Λ^* is given by (6-64) and $\chi^2_{(g-1)(b-1)p}(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with $(g - 1)(b - 1)p$ d.f.

Ordinarily, the test for interaction is carried out before the tests for main factor effects. If interaction effects exist, the factor effects do not have a clear interpretation. From a practical standpoint, it is not advisable to proceed with the additional multivariate tests. Instead, p univariate two-way analyses of variance (one for each variable) are often conducted to see whether the interaction appears in some responses but not

[5]The likelihood test procedures require that $p \leq gb(n - 1)$, so that $\mathbf{SSP}_{res}$ will be positive definite (with probability 1).

others. Those responses without interaction may be interpreted in terms of additive factor 1 and 2 effects, provided that the latter effects exist. In any event, interaction plots similar to Figure 6.3, but with treatment sample means replacing expected values, best clarify the relative magnitudes of the main and interaction effects.

In the multivariate model, we test for factor 1 and factor 2 main effects as follows. First, consider the hypotheses H_0: $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = \mathbf{0}$ and H_1: at least one $\boldsymbol{\tau}_\ell \neq \mathbf{0}$. These hypotheses specify *no* factor 1 effects and *some* factor 1 effects, respectively. Let

$$\Lambda^* = \frac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{fac1}} + \text{SSP}_{\text{res}}|} \tag{6-66}$$

so that small values of Λ^* are consistent with H_1. Using Bartlett's correction, the likelihood ratio test is as follows:

Reject H_0: $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = \mathbf{0}$ (no factor 1 effects) at level α if

$$-\left[gb(n-1) - \frac{p+1-(g-1)}{2}\right] \ln \Lambda^* > \chi^2_{(g-1)p}(\alpha) \tag{6-67}$$

where Λ^* is given by (6-66) and $\chi^2_{(g-1)p}(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with $(g-1)p$ d.f.

In a similar manner, factor 2 effects are tested by considering H_0: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_b = \mathbf{0}$ and H_1: at least one $\boldsymbol{\beta}_k \neq \mathbf{0}$. Small values of

$$\Lambda^* = \frac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{fac2}} + \text{SSP}_{\text{res}}|} \tag{6-68}$$

are consistent with H_1. Once again, for large samples and using Bartlett's correction: Reject H_0: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_b = \mathbf{0}$ (no factor 2 effects) at level α if

$$-\left[gb(n-1) - \frac{p+1-(b-1)}{2}\right] \ln \Lambda^* > \chi^2_{(b-1)p}(\alpha) \tag{6-69}$$

where Λ^* is given by (6-68) and $\chi^2_{(b-1)p}(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with $(b-1)p$ degrees of freedom.

Simultaneous confidence intervals for contrasts in the model parameters can provide insights into the nature of the factor effects. Results comparable to Result 6.5 are available for the two-way model. When interaction effects are negligible, we may concentrate on contrasts in the factor 1 and factor 2 main effects. The Bonferroni approach applies to the components of the differences $\boldsymbol{\tau}_\ell - \boldsymbol{\tau}_m$ of the factor 1 effects and the components of $\boldsymbol{\beta}_k - \boldsymbol{\beta}_q$ of the factor 2 effects, respectively.

The $100(1-\alpha)\%$ simultaneous confidence intervals for $\tau_{\ell i} - \tau_{mi}$ are

$$\tau_{\ell i} - \tau_{mi} \quad \text{belongs to} \quad (\bar{x}_{\ell \cdot i} - \bar{x}_{m \cdot i}) \pm t_\nu\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{E_{ii}}{\nu}\frac{2}{bn}} \tag{6-70}$$

where $\nu = gb(n-1)$, E_{ii} is the ith diagonal element of $\mathbf{E} = \text{SSP}_{\text{res}}$, and $\bar{x}_{\ell \cdot i} - \bar{x}_{m \cdot i}$ is the ith component of $\bar{\mathbf{x}}_{\ell \cdot} - \bar{\mathbf{x}}_{m \cdot}$.

Similarly, the $100(1 - \alpha)$ percent simultaneous confidence intervals for $\beta_{ki} - \beta_{qi}$ are

$$\beta_{ki} - \beta_{qi} \quad \text{belongs to} \quad (\bar{x}_{\cdot ki} - \bar{x}_{\cdot qi}) \pm t_{\nu}\left(\frac{\alpha}{pb(b-1)}\right)\sqrt{\frac{E_{ii}}{\nu}\frac{2}{gn}} \qquad (6\text{-}71)$$

where ν and E_{ii} are as just defined and $\bar{x}_{\cdot ki} - \bar{x}_{\cdot qi}$ is the ith component of $\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}}_{\cdot q}$.

Comment. We have considered the multivariate two-way model with replications. That is, the model allows for n replications of the responses at each combination of factor levels. This enables us to examine the "interaction" of the factors. If only one observation vector is available at each combination of factor levels, the two-way model does not allow for the possibility of a general interaction term $\gamma_{\ell k}$. The corresponding MANOVA table includes only factor 1, factor 2, and residual sources of variation as components of the total variation. (See Exercise 6.13.)

Example 6.13 (A two-way multivariate analysis of variance of plastic film data) The optimum conditions for extruding plastic film have been examined using a technique called Evolutionary Operation. (See [9].) In the course of the study that was done, three responses—X_1 = tear resistance, X_2 = gloss, and X_3 = opacity—were measured at two levels of the factors, *rate of extrusion* and *amount of an additive*. The measurements were repeated $n = 5$ times at each combination of the factor levels. The data are displayed in Table 6.4.

Table 6.4 Plastic Film Data

x_1 = tear resistance, x_2 = gloss, and x_3 = opacity

		Factor 2: Amount of additive					
		Low (1.0%)			High (1.5%)		
		x_1	x_2	x_3	x_1	x_2	x_3
Factor 1: Change in rate of extrusion	Low (−10)%	[6.5	9.5	4.4]	[6.9	9.1	5.7]
		[6.2	9.9	6.4]	[7.2	10.0	2.0]
		[5.8	9.6	3.0]	[6.9	9.9	3.9]
		[6.5	9.6	4.1]	[6.1	9.5	1.9]
		[6.5	9.2	0.8]	[6.3	9.4	5.7]
		x_1	x_2	x_3	x_1	x_2	x_3
	High (10%)	[6.7	9.1	2.8]	[7.1	9.2	8.4]
		[6.6	9.3	4.1]	[7.0	8.8	5.2]
		[7.2	8.3	3.8]	[7.2	9.7	6.9]
		[7.1	8.4	1.6]	[7.5	10.1	2.7]
		[6.8	8.5	3.4]	[7.6	9.2	1.9]

The matrices of the appropriate sum of squares and cross products were calculated (see the SAS statistical software output in Panel 6.1[6]), leading to the following MANOVA table:

[6]Additional SAS programs for MANOVA and other procedures discussed in this chapter are available in [13].

Source of variation	SSP	d.f.
Factor 1: change in rate of extrusion	$\begin{bmatrix} 1.7405 & -1.5045 & .8555 \\ & 1.3005 & -.7395 \\ & & .4205 \end{bmatrix}$	1
Factor 2: amount of additive	$\begin{bmatrix} .7605 & .6825 & 1.9305 \\ & .6125 & 1.7325 \\ & & 4.9005 \end{bmatrix}$	1
Interaction	$\begin{bmatrix} .0005 & .0165 & .0445 \\ & .5445 & 1.4685 \\ & & 3.9605 \end{bmatrix}$	1
Residual	$\begin{bmatrix} 1.7640 & .0200 & -3.0700 \\ & 2.6280 & -.5520 \\ & & 64.9240 \end{bmatrix}$	16
Total (corrected)	$\begin{bmatrix} 4.2655 & -.7855 & -.2395 \\ & 5.0855 & 1.9095 \\ & & 74.2055 \end{bmatrix}$	19

PANEL 6.1 SAS ANALYSIS FOR EXAMPLE 6.13 USING PROC GLM

```
title 'MANOVA';
data film;
infile 'T6-4.dat';
input x1 x2 x3 factor1 factor2;
proc glm data = film;
class factor1 factor2;
model x1 x2 x3 = factor1 factor2 factor1*factor2 /ss3;
manova h = factor1 factor2 factor1*factor2 /printe;
means factor1 factor2;
```

PROGRAM COMMANDS

General Linear Models Procedure
Class Level Information

OUTPUT

Class	Levels	Values
FACTOR1	2	0 1
FACTOR2	2	0 1

Number of observations in data set = 20

Dependent Variable: X1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2.50150000	0.83383333	7.56	0.0023
Error	16	1.76400000	0.11025000		
Corrected Total	19	4.26550000			

R-Square	C.V.	Root MSE	X1 Mean
0.586449	4.893724	0.332039	6.78500000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
FACTOR1	1	1.74050000	1.74050000	15.79	0.0011
FACTOR2	1	0.76050000	0.76050000	6.90	0.0183
FACTOR1*FACTOR2	1	0.00050000	0.00050000	0.00	0.9471

(continues on next page)

PANEL 6.1 *(continued)*

Dependent Variable: X2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2.45750000	0.81916667	4.99	0.0125
Error	16	2.62800000	0.16425000		
Corrected Total	19	5.08550000			

R-Square	C.V.	Root MSE	X2 Mean
0.483237	4.350807	0.405278	9.31500000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
FACTOR1	1	1.30050000	1.30050000	7.92	0.0125
FACTOR2	1	0.61250000	0.61250000	3.73	0.0714
FACTOR1*FACTOR2	1	0.54450000	0.54450000	3.32	0.0874

Dependent Variable: X3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	9.28150000	3.09383333	0.76	0.5315
Error	16	64.92400000	4.05775000		
Corrected Total	19	74.20550000			

R-Square	C.V.	Root MSE	X3 Mean
0.125078	51.19151	2.014386	3.93500000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
FACTOR	1	0.42050000	0.42050000	0.10	0.7517
FACTOR2	1	4.90050000	4.90050000	1.21	0.2881
FACTOR1*FACTOR2	1	3.96050000	3.96050000	0.98	0.3379

E = Error SS&CP Matrix

	X1	X2	X3
X1	1.764	0.02	-3.07
X2	0.02	2.628	-0.552
X3	-3.07	-0.552	64.924

Manova Test Criteria and Exact F Statistics for

the **Hypothesis of no Overall FACTOR1 Effect**

H = Type III SS&CP Matrix for FACTOR1 E = Error SS&CP Matrix

S = 1 M = 0.5 N = 6

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	**0.38185838**	**7.5543**	**3**	**14**	**0.0030**
Pillai's Trace	0.61814162	7.5543	3	14	0.0030
Hotelling-Lawley Trace	1.61877188	7.5543	3	14	0.0030
Roy's Greatest Root	1.61877188	7.5543	3	14	0.0030

(continues on next page)

PANEL 6.1 *(continued)*

Manova Test Criteria and Exact F Statistics for

the **Hypothesis of no Overall FACTOR2 Effect**

H = Type III SS&CP Matrix for FACTOR2 E = Error SS&CP Matrix
S = 1 M = 0.5 N = 6

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	**0.52303490**	**4.2556**	**3**	**14**	**0.0247**
Pillai's Trace	0.47696510	4.2556	3	14	0.0247
Hotelling-Lawley Trace	0.91191832	4.2556	3	14	0.0247
Roy's Greatest Root	0.91191832	4.2556	3	14	0.0247

Manova Test Criteria and Exact F Statistics for

the **Hypothesis of no Overall FACTOR1*FACTOR2 Effect**

H = Type III SS&CP Matrix for FACTOR1*FACTOR2 E = Error SS&CP Matrix
S = 1 M = 0.5 N = 6

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	**0.77710576**	**1.3385**	**3**	**14**	**0.3018**
Pillai's Trace	0.22289424	1.3385	3	14	0.3018
Hotelling-Lawley Trace	0.28682614	1.3385	3	14	0.3018
Roy's Greatest Root	0.28682614	1.3385	3	14	0.3018

Level of		---------X1---------		---------X2---------	
FACTOR1	N	Mean	SD	Mean	SD
0	10	6.49000000	0.42018514	9.57000000	0.29832868
1	10	7.08000000	0.32249031	9.06000000	0.57580861

Level of		---------X3---------	
FACTOR1	N	Mean	SD
0	10	3.79000000	1.85379491
1	10	4.08000000	2.18214981

Level of		---------X1---------		---------X2---------	
FACTOR2	N	Mean	SD	Mean	SD
0	10	6.59000000	0.40674863	9.14000000	0.56015871
1	10	6.98000000	0.47328638	9.49000000	0.42804465

Level of		---------X3---------	
FACTOR2	N	Mean	SD
0	10	3.44000000	1.55077042
1	10	4.43000000	2.30123155

To test for interaction, we compute

$$\Lambda^* = \frac{|\mathrm{SSP}_{\mathrm{res}}|}{|\mathrm{SSP}_{\mathrm{int}} + \mathrm{SSP}_{\mathrm{res}}|} = \frac{275.7098}{354.7906} = .7771$$

For $(g - 1)(b - 1) = 1$,

$$F = \left(\frac{1 - \Lambda^*}{\Lambda^*}\right)\frac{(gb(n-1) - p + 1)/2}{(|(g-1)(b-1) - p| + 1)/2}$$

has an exact F-distribution with $\nu_1 = |(g - 1)(b - 1) - p| + 1$ and ν_2 $gb(n - 1) - p + 1$ d.f. (See [1].) For our example.

$$F = \left(\frac{1 - .7771}{.7771}\right)\frac{(2(2)(4) - 3 + 1)/2}{(|1(1) - 3| + 1)/2} = 1.34$$

$$\nu_1 = (|1(1) - 3| + 1) = 3$$

$$\nu_2 = (2(2)(4) - 3 + 1) = 14$$

and $F_{3,14}(.05) = 3.34$. Since $F = 1.34 < F_{3,14}(.05) = 3.34$, we do not reject the hypothesis $H_0\colon \boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = \boldsymbol{\gamma}_{21} = \boldsymbol{\gamma}_{22} = \mathbf{0}$ (no interaction effects).

Note that the approximate chi-square statistic for this test is $-[2(2)(4) - (3 + 1 - 1(1))/2] \ln(.7771) = 3.66$, from (6-65). Since $\chi_3^2(.05) = 7.81$, we would reach the same conclusion as provided by the exact F-test.

To test for factor 1 and factor 2 effects (see page 317), we calculate

$$\Lambda_1^* = \frac{|\mathrm{SSP}_{\mathrm{res}}|}{|\mathrm{SSP}_{\mathrm{fac}\,1} + \mathrm{SSP}_{\mathrm{res}}|} = \frac{275.7098}{722.0212} = .3819$$

and

$$\Lambda_2^* = \frac{|\mathrm{SSP}_{\mathrm{res}}|}{|\mathrm{SSP}_{\mathrm{fac}\,2} + \mathrm{SSP}_{\mathrm{res}}|} = \frac{275.7098}{527.1347} = .5230$$

For both $g - 1 = 1$ and $b - 1 = 1$,

$$F_1 = \left(\frac{1 - \Lambda_1^*}{\Lambda_1^*}\right)\frac{(gb(n-1) - p + 1)/2}{(|(g-1) - p| + 1)/2}$$

and

$$F_2 = \left(\frac{1 - \Lambda_2^*}{\Lambda_2^*}\right)\frac{(gb(n-1) - p + 1)/2}{(|(b-1) - p| + 1)/2}$$

have F-distributions with degrees of freedom $\nu_1 = |(g - 1) - p| + 1$, $\nu_2 = gb(n - 1) - p + 1$ and $\nu_1 = |(b - 1) - p| + 1$, $\nu_2 = gb(n - 1) - p + 1$, respectively. (See [1].) In our case,

$$F_1 = \left(\frac{1 - .3819}{.3819}\right)\frac{(16 - 3 + 1)/2}{(|1 - 3| + 1)/2} = 7.55$$

$$F_2 = \left(\frac{1 - .5230}{.5230}\right)\frac{(16 - 3 + 1)/2}{(|1 - 3| + 1)/2} = 4.26$$

and

$$\nu_1 = |1 - 3| + 1 = 3 \qquad \nu_2 = (16 - 3 + 1) = 14$$

From before, $F_{3,14}(.05) = 3.34$. We have $F_1 = 7.55 > F_{3,14}(.05) = 3.34$, and therefore, we reject $H_0: \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \mathbf{0}$ (no factor 1 effects) at the 5% level. Similarly, $F_2 = 4.26 > F_{3,14}(.05) = 3.34$, and we reject $H_0: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \mathbf{0}$ (no factor 2 effects) at the 5% level. We conclude that both the *change in rate of extrusion* and the *amount of additive* affect the responses, and they do so in an additive manner.

The *nature* of the effects of factors 1 and 2 on the responses is explored in Exercise 6.15. In that exercise, simultaneous confidence intervals for contrasts in the components of $\boldsymbol{\tau}_\ell$ and $\boldsymbol{\beta}_k$ are considered. ■

6.8 Profile Analysis

Profile analysis pertains to situations in which a battery of p treatments (tests, questions, and so forth) are administered to two or more groups of subjects. All responses must be expressed in similar units. Further, it is assumed that the responses for the different groups are independent of one another. Ordinarily, we might pose the question, are the population mean vectors the same? In profile analysis, the question of equality of mean vectors is divided into several specific possibilities.

Consider the population means $\boldsymbol{\mu}_1' = [\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14}]$ representing the average responses to four treatments for the first group. A plot of these means, connected by straight lines, is shown in Figure 6.4. This broken-line graph is the *profile* for population 1.

Profiles can be constructed for each population (group). We shall concentrate on two groups. Let $\boldsymbol{\mu}_1' = [\mu_{11}, \mu_{12}, \ldots, \mu_{1p}]$ and $\boldsymbol{\mu}_2' = [\mu_{21}, \mu_{22}, \ldots, \mu_{2p}]$ be the mean responses to p treatments for populations 1 and 2, respectively. The hypothesis $H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ implies that the treatments have the same (average) effect on the two populations. In terms of the population profiles, we can formulate the question of equality in a stepwise fashion.

1. Are the profiles parallel?
   Equivalently: Is $H_{01}: \mu_{1i} - \mu_{1i-1} = \mu_{2i} - \mu_{2i-1}$, $i = 2, 3, \ldots, p$, acceptable?
2. Assuming that the profiles *are* parallel, are the profiles coincident?[7]
   Equivalently: Is $H_{02}: \mu_{1i} = \mu_{2i}$, $i = 1, 2, \ldots, p$, acceptable?

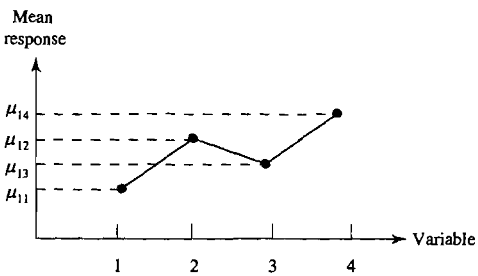


Figure 6.4 The population profile $p = 4$.

[7]The question, "Assuming that the profiles are parallel, are the profiles linear?" is considered in Exercise 6.12. The null hypothesis of parallel linear profiles can be written $H_0: (\mu_{1i} + \mu_{2i}) - (\mu_{1i-1} + \mu_{2i-1}) = (\mu_{1i-1} + \mu_{2i-1}) - (\mu_{1i-2} + \mu_{2i-2})$, $i = 3, \ldots, p$. Although this hypothesis may be of interest in a particular situation, in practice the question of whether two parallel profiles are the same (coincident), whatever their nature, is usually of greater interest.

3. Assuming that the profiles *are* coincident, are the profiles level? That is, are all the means equal to the same constant?
   Equivalently: Is $H_{03}: \mu_{11} = \mu_{12} = \cdots = \mu_{1p} = \mu_{21} = \mu_{22} = \cdots = \mu_{2p}$ acceptable?

   The null hypothesis in stage 1 can be written

$$H_{01}: \mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2$$

where $\mathbf{C}$ is the contrast matrix

$$\underset{((p-1)\times p)}{\mathbf{C}} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \tag{6-72}$$

For independent samples of sizes n_1 and n_2 from the two populations, the null hypothesis can be tested by constructing the transformed observations

$$\mathbf{C}\mathbf{x}_{1j}, \qquad j = 1, 2, \ldots, n_1$$

and

$$\mathbf{C}\mathbf{x}_{2j}, \qquad j = 1, 2, \ldots, n_2$$

These have sample mean vectors $\mathbf{C}\bar{\mathbf{x}}_1$ and $\mathbf{C}\bar{\mathbf{x}}_2$, respectively, and pooled covariance matrix $\mathbf{C}\mathbf{S}_{\text{pooled}}\mathbf{C}'$.

Since the two sets of transformed observations have $N_{p-1}(\mathbf{C}\boldsymbol{\mu}_1, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$ and $N_{p-1}(\mathbf{C}\boldsymbol{\mu}_2, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$ distributions, respectively, an application of Result 6.2 provides a test for parallel profiles.

Test for Parallel Profiles for Two Normal Populations

Reject $H_{01}: \mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2$ (parallel profiles) at level α if

$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{C}'\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{C}\mathbf{S}_{\text{pooled}}\mathbf{C}'\right]^{-1}\mathbf{C}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) > c^2 \tag{6-73}$$

where

$$c^2 = \frac{(n_1 + n_2 - 2)(p - 1)}{n_1 + n_2 - p} F_{p-1, n_1+n_2-p}(\alpha)$$

When the profiles are parallel, the first is either above the second ($\mu_{1i} > \mu_{2i}$, for all i), or vice versa. Under this condition, the profiles will be coincident only if the total heights $\mu_{11} + \mu_{12} + \cdots + \mu_{1p} = \mathbf{1}'\boldsymbol{\mu}_1$ and $\mu_{21} + \mu_{22} + \cdots + \mu_{2p} = \mathbf{1}'\boldsymbol{\mu}_2$ are equal. Therefore, the null hypothesis at stage 2 can be written in the equivalent form

$$H_{02}: \mathbf{1}'\boldsymbol{\mu}_1 = \mathbf{1}'\boldsymbol{\mu}_2$$

We can then test H_{02} with the usual two-sample t-statistic based on the univariate observations $\mathbf{1}'\mathbf{x}_{1j}, j = 1, 2, \ldots, n_1$, and $\mathbf{1}'\mathbf{x}_{2j}, j = 1, 2, \ldots, n_2$.

Test for Coincident Profiles, Given That Profiles Are Parallel

For two normal populations, reject H_{02}: $\mathbf{1}'\boldsymbol{\mu}_1 = \mathbf{1}'\boldsymbol{\mu}_2$ (profiles coincident) at level α if

$$T^2 = \mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{1}'\mathbf{S}_{\text{pooled}}\mathbf{1}\right]^{-1}\mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

$$= \left(\frac{\mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{1}'\mathbf{S}_{\text{pooled}}\mathbf{1}}}\right)^2 > t^2_{n_1+n_2-2}\left(\frac{\alpha}{2}\right) = F_{1,n_1+n_2-2}(\alpha) \qquad (6\text{-}74)$$

For coincident profiles, $\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1n_1}$ and $\mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2n_2}$ are all observations from the same normal population? The next step is to see whether all variables have the same mean, so that the common profile is level.

When H_{01} and H_{02} are tenable, the common mean vector $\boldsymbol{\mu}$ is estimated, using all $n_1 + n_2$ observations, by

$$\bar{\mathbf{x}} = \frac{1}{n_1 + n_2}\left(\sum_{j=1}^{n_1}\mathbf{x}_{1j} + \sum_{j=1}^{n_2}\mathbf{x}_{2j}\right) = \frac{n_1}{(n_1 + n_2)}\bar{\mathbf{x}}_1 + \frac{n_2}{(n_1 + n_2)}\bar{\mathbf{x}}_2$$

If the common profile is level, then $\mu_1 = \mu_2 = \cdots = \mu_p$, and the null hypothesis at stage 3 can be written as

$$H_{03}: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$$

where $\mathbf{C}$ is given by (6-72). Consequently, we have the following test.

Test for Level Profiles, Given That Profiles Are Coincident

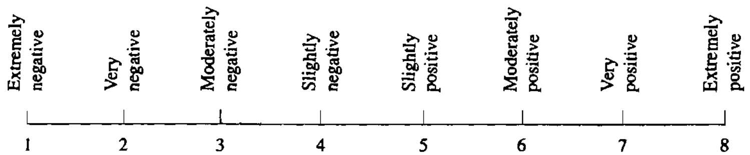
For two normal populations: Reject H_{03}: $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ (profiles level) at level α if

$$(n_1 + n_2)\bar{\mathbf{x}}'\mathbf{C}'[\mathbf{C}\mathbf{S}\mathbf{C}']^{-1}\mathbf{C}\bar{\mathbf{x}} > c^2 \qquad (6\text{-}75)$$

where $\mathbf{S}$ is the sample covariance matrix based on all $n_1 + n_2$ observations and

$$c^2 = \frac{(n_1 + n_2 - 1)(p - 1)}{(n_1 + n_2 - p + 1)}F_{p-1,n_1+n_2-p+1}(\alpha)$$

Example 6.14 (A profile analysis of love and marriage data) As part of a larger study of love and marriage, E. Hatfield, a sociologist, surveyed adults with respect to their marriage "contributions" and "outcomes" and their levels of "passionate" and "companionate" love. Recently married males and females were asked to respond to the following questions, using the 8-point scale in the figure below.



1. All things considered, how would you describe *your contributions* to the marriage?
2. All things considered, how would you describe *your outcomes* from the marriage?

Subjects were also asked to respond to the following questions, using the 5-point scale shown.

3. What is the level of *passionate* love that you feel for your partner?
4. What is the level of *companionate* love that you feel for your partner?

None at all	Very little	Some	A great deal	Tremendous amount
1	2	3	4	5

Let

$$\begin{aligned} x_1 &= \text{an 8-point scale response to Question 1} \\ x_2 &= \text{an 8-point scale response to Question 2} \\ x_3 &= \text{a 5-point scale response to Question 3} \\ x_4 &= \text{a 5-point scale response to Question 4} \end{aligned}$$

and the two populations be defined as

$$\begin{aligned} \text{Population 1} &= \text{married men} \\ \text{Population 2} &= \text{married women} \end{aligned}$$

The population means are the average responses to the $p = 4$ questions for the populations of males and females. Assuming a common covariance matrix $\boldsymbol{\Sigma}$, it is of interest to see whether the profiles of males and females are the same.

A sample of $n_1 = 30$ males and $n_2 = 30$ females gave the sample mean vectors

$$\bar{\mathbf{x}}_1 = \underset{\text{(males)}}{\begin{bmatrix} 6.833 \\ 7.033 \\ 3.967 \\ 4.700 \end{bmatrix}}, \qquad \bar{\mathbf{x}}_2 = \underset{\text{(females)}}{\begin{bmatrix} 6.633 \\ 7.000 \\ 4.000 \\ 4.533 \end{bmatrix}}$$

and pooled covariance matrix

$$\mathbf{S}_{\text{pooled}} = \begin{bmatrix} .606 & .262 & .066 & .161 \\ .262 & .637 & .173 & .143 \\ .066 & .173 & .810 & .029 \\ .161 & .143 & .029 & .306 \end{bmatrix}$$

The sample mean vectors are plotted as sample profiles in Figure 6.5 on page 327.

Since the sample sizes are reasonably large, we shall use the normal theory methodology, even though the data, which are integers, are clearly nonnormal. To test for parallelism (H_{01}: $\mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2$), we compute

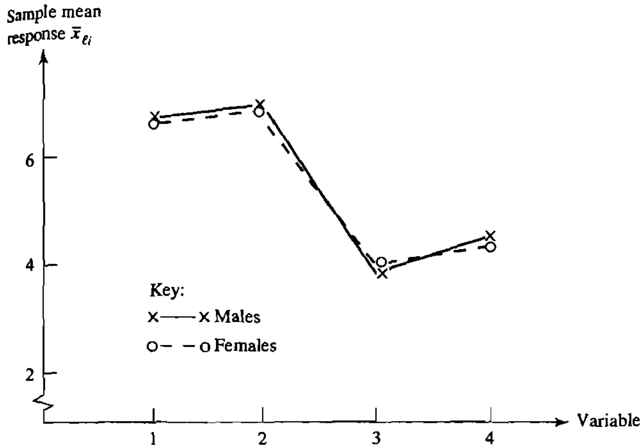


Figure 6.5 Sample profiles for marriage–love responses.

$$\mathbf{C}\mathbf{S}_{\text{pooled}}\mathbf{C}' = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \mathbf{S}_{\text{pooled}} \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} .719 & -.268 & -.125 \\ -.268 & 1.101 & -.751 \\ -.125 & -.751 & 1.058 \end{bmatrix}$$

and

$$\mathbf{C}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} .200 \\ .033 \\ -.033 \\ .167 \end{bmatrix} = \begin{bmatrix} -.167 \\ -.066 \\ .200 \end{bmatrix}$$

Thus,

$$T^2 = [-.167, -.066, .200]\left(\tfrac{1}{30} + \tfrac{1}{30}\right)^{-1} \begin{bmatrix} .719 & -.268 & -.125 \\ -.268 & 1.101 & -.751 \\ -.125 & -.751 & 1.058 \end{bmatrix}^{-1} \begin{bmatrix} -.167 \\ -.066 \\ .200 \end{bmatrix}$$
$$= 15(.067) = 1.005$$

Moreover, with $\alpha = .05$, $c^2 = [(30+30-2)(4-1)/(30+30-4)]F_{3,56}(.05) = 3.11(2.8) = 8.7$. Since $T^2 = 1.005 < 8.7$, we conclude that the hypothesis of parallel profiles for men and women is tenable. Given the plot in Figure 6.5, this finding is not surprising.

Assuming that the profiles are parallel, we can test for *coincident* profiles. To test H_{02}: $\mathbf{1}'\boldsymbol{\mu}_1 = \mathbf{1}'\boldsymbol{\mu}_2$ (profiles coincident), we need

$$\text{Sum of elements in } (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = \mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = .367$$
$$\text{Sum of elements in } \mathbf{S}_{\text{pooled}} = \mathbf{1}'\mathbf{S}_{\text{pooled}}\mathbf{1} = 4.207$$

Using (6-74), we obtain

$$T^2 = \left(\frac{.367}{\sqrt{\left(\frac{1}{30} + \frac{1}{30}\right)4.027}} \right)^2 = .501$$

With $\alpha = .05$, $F_{1,58}(.05) = 4.0$, and $T^2 = .501 < F_{1,58}(.05) = 4.0$, we cannot reject the hypothesis that the profiles are coincident. That is, the responses of men and women to the four questions posed appear to be the same.

We could now test for level profiles; however, it does not make sense to carry out this test for our example, since Questions 1 and 2 were measured on a scale of 1–8, while Questions 3 and 4 were measured on a scale of 1–5. The incompatibility of these scales makes the test for level profiles meaningless and illustrates the need for similar measurements in order to carry out a complete profile analysis. ■

When the sample sizes are small, a profile analysis will depend on the normality assumption. This assumption can be checked, using methods discussed in Chapter 4, with the original observations $\mathbf{x}_{\ell j}$ or the contrast observations $\mathbf{C}\mathbf{x}_{\ell j}$.

The analysis of profiles for several populations proceeds in much the same fashion as that for two populations. In fact, the general measures of comparison are analogous to those just discussed. (See [13], [18].)

6.9 Repeated Measures Designs and Growth Curves

As we said earlier, the term "repeated measures" refers to situations where the same characteristic is observed, at different times or locations, on the same subject.

(a) The observations on a subject may correspond to different treatments as in Example 6.2 where the time between heartbeats was measured under the 2×2 treatment combinations applied to each dog. The treatments need to be compared when the responses on the same subject are correlated.

(b) A single treatment may be applied to each subject and a single characteristic observed over a period of time. For instance, we could measure the weight of a puppy at birth and then once a month. It is the curve traced by a typical dog that must be modeled. In this context, we refer to the curve as a *growth curve*.

When some subjects receive one treatment and others another treatment, the growth curves for the treatments need to be compared.

To illustrate the growth curve model introduced by Potthoff and Roy [21], we consider calcium measurements of the dominant ulna bone in older women. Besides an initial reading, Table 6.5 gives readings after one year, two years, and three years for the control group. Readings obtained by photon absorptiometry from the same subject are correlated but those from different subjects should be independent. The model assumes that the same covariance matrix $\mathbf{\Sigma}$ holds for each subject. Unlike univariate approaches, this model does not require the four measurements to have equal variances. A profile, constructed from the four sample means $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4)$, summarizes the growth which here is a loss of calcium over time. Can the growth pattern be adequately represented by a polynomial in time?

Table 6.5 Calcium Measurements on the Dominant Ulna; Control Group

Subject	Initial	1 year	2 year	3 year
1	87.3	86.9	86.7	75.5
2	59.0	60.2	60.0	53.6
3	76.7	76.5	75.7	69.5
4	70.6	76.1	72.1	65.3
5	54.9	55.1	57.2	49.0
6	78.2	75.3	69.1	67.6
7	73.7	70.8	71.8	74.6
8	61.8	68.7	68.2	57.4
9	85.3	84.4	79.2	67.0
10	82.3	86.9	79.4	77.4
11	68.6	65.4	72.3	60.8
12	67.8	69.2	66.3	57.9
13	66.2	67.0	67.0	56.2
14	81.0	82.3	86.8	73.9
15	72.3	74.6	75.3	66.1
Mean	72.38	73.29	72.47	64.79

Source: Data courtesy of Everett Smith.

When the p measurements on all subjects are taken at times $t_1, t_2, \ldots, t_p$, the Potthoff–Roy model for quadratic growth becomes

$$E[\mathbf{X}] = E\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 t_1 + \beta_2 t_1^2 \\ \beta_0 + \beta_1 t_2 + \beta_2 t_2^2 \\ \vdots \\ \beta_0 + \beta_1 t_p + \beta_2 t_p^2 \end{bmatrix}$$

where the ith mean μ_i is the quadratic expression evaluated at t_i.

Usually groups need to be compared. Table 6.6 gives the calcium measurements for a second set of women, the treatment group, that received special help with diet and a regular exercise program.

When a study involves several treatment groups, an extra subscript is needed as in the one-way MANOVA model. Let $\mathbf{X}_{\ell 1}, \mathbf{X}_{\ell 2}, \ldots, \mathbf{X}_{\ell n_\ell}$ be the n_ℓ vectors of measurements on the n_ℓ subjects in group ℓ, for $\ell = 1, \ldots, g$.

Assumptions. All of the $\mathbf{X}_{\ell j}$ are independent and have the same covariance matrix $\mathbf{\Sigma}$. Under the quadratic growth model, the mean vectors are

$$E[\mathbf{X}_{\ell j}] = \begin{bmatrix} \beta_{\ell 0} + \beta_{\ell 1} t_1 + \beta_{\ell 2} t_1^2 \\ \beta_{\ell 0} + \beta_{\ell 1} t_2 + \beta_{\ell 2} t_2^2 \\ \vdots \\ \beta_{\ell 0} + \beta_{\ell 1} t_p + \beta_{\ell 2} t_p^2 \end{bmatrix} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_p & t_p^2 \end{bmatrix} \begin{bmatrix} \beta_{\ell 0} \\ \beta_{\ell 1} \\ \beta_{\ell 2} \end{bmatrix} = \mathbf{B}\boldsymbol{\beta}_\ell$$

Table 6.6 Calcium Measurements on the Dominant Ulna; Treatment Group

Subject	Initial	1 year	2 year	3 year
1	83.8	85.5	86.2	81.2
2	65.3	66.9	67.0	60.6
3	81.2	79.5	84.5	75.2
4	75.4	76.7	74.3	66.7
5	55.3	58.3	59.1	54.2
6	70.3	72.3	70.6	68.6
7	76.5	79.9	80.4	71.6
8	66.0	70.9	70.3	64.1
9	76.7	79.0	76.9	70.3
10	77.2	74.0	77.8	67.9
11	67.3	70.7	68.9	65.9
12	50.3	51.4	53.6	48.0
13	57.7	57.0	57.5	51.5
14	74.3	77.7	72.6	68.0
15	74.0	74.7	74.5	65.7
16	57.3	56.0	64.7	53.0
Mean	69.29	70.66	71.18	64.53

Source: Data courtesy of Everett Smith.

where

$$\mathbf{B} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_p & t_p^2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta}_\ell = \begin{bmatrix} \beta_{\ell 0} \\ \beta_{\ell 1} \\ \beta_{\ell 2} \end{bmatrix} \tag{6-76}$$

If a qth-order polynomial is fit to the growth data, then

$$\mathbf{B} = \begin{bmatrix} 1 & t_1 & \cdots & t_1^q \\ 1 & t_2 & \cdots & t_2^q \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & t_p & \cdots & t_p^q \end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta}_\ell = \begin{bmatrix} \beta_{\ell 0} \\ \beta_{\ell 1} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{\ell q} \end{bmatrix} \tag{6-77}$$

Under the assumption of multivariate normality, the maximum likelihood estimators of the $\boldsymbol{\beta}_\ell$ are

$$\hat{\boldsymbol{\beta}}_\ell = (\mathbf{B}'\mathbf{S}_{\text{pooled}}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{S}_{\text{pooled}}^{-1}\overline{\mathbf{X}}_\ell \quad \text{for} \quad \ell = 1, 2, \ldots, g \tag{6-78}$$

where

$$\mathbf{S}_{\text{pooled}} = \frac{1}{(N - g)}((n_1 - 1)\mathbf{S}_1 + \cdots + (n_g - 1)\mathbf{S}_g) = \frac{1}{N - g}\mathbf{W}$$

with $N = \sum_{\ell=1}^{g} n_\ell$, is the pooled estimator of the common covariance matrix $\boldsymbol{\Sigma}$. The estimated covariances of the maximum likelihood estimators are

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}_\ell) = \frac{k}{n_\ell}(\mathbf{B}'\mathbf{S}_{\text{pooled}}^{-1}\mathbf{B})^{-1} \quad \text{for} \quad \ell = 1, 2, \ldots, g \tag{6-79}$$

where $k = (N - g)(N - g - 1)/(N - g - p + q)(N - g - p + q + 1)$.

Also, $\hat{\boldsymbol{\beta}}_\ell$ and $\hat{\boldsymbol{\beta}}_h$ are independent, for $\ell \neq h$, so their covariance is $\mathbf{0}$.

We can formally test that a qth-order polynomial is adequate. The model is fit without restrictions, the error sum of squares and cross products matrix is just the within groups $\mathbf{W}$ that has $N - g$ degrees of freedom. Under a qth-order polynomial, the error sum of squares and cross products

$$\mathbf{W}_q = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{X}_{\ell j} - \mathbf{B}\hat{\boldsymbol{\beta}}_\ell)(\mathbf{X}_{\ell j} - \mathbf{B}\hat{\boldsymbol{\beta}}_\ell)' \tag{6-80}$$

has $n_g - g + p - q - 1$ degrees of freedom. The likelihood ratio test of the null hypothesis that the q-order polynomial is adequate can be based on Wilks' lambda

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{W}_q|} \tag{6-81}$$

Under the polynomial growth model, there are $q + 1$ terms instead of the p means for each of the groups. Thus there are $(p - q - 1)g$ fewer parameters. For large sample sizes, the null hypothesis that the polynomial is adequate is rejected if

$$-\left(N - \frac{1}{2}(p - q + g)\right)\ln \Lambda^* > \chi^2_{(p-q-1)g}(\alpha) \tag{6-82}$$

Example 6.15 (Fitting a quadratic growth curve to calcium loss) Refer to the data in Tables 6.5 and 6.6. Fit the model for quadratic growth.

A computer calculation gives

$$[\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2] = \begin{bmatrix} 73.0701 & 70.1387 \\ 3.6444 & 4.0900 \\ -2.0274 & -1.8534 \end{bmatrix}$$

so the estimated growth curves are

$$\begin{array}{llll} \text{Control group:} & 73.07 & +\ 3.64t & -\ 2.03t^2 \\ & (2.58) & (.83) & (.28) \\ \text{Treatment group:} & 70.14 & +\ 4.09t & -\ 1.85t^2 \\ & (2.50) & (.80) & (.27) \end{array}$$

where

$$(\mathbf{B}'\mathbf{S}_{\text{pooled}}^{-1}\mathbf{B})^{-1} = \begin{bmatrix} 93.1744 & -5.8368 & 0.2184 \\ -5.8368 & 9.5699 & -3.0240 \\ 0.2184 & -3.0240 & 1.1051 \end{bmatrix}$$

and, by (6-79), the standard errors given below the parameter estimates were obtained by dividing the diagonal elements by n_ℓ and taking the square root.

Examination of the estimates and the standard errors reveals that the t^2 terms are needed. Loss of calcium is predicted after 3 years for both groups. Further, there does not seem to be any substantial difference between the two groups.

Wilks' lambda for testing the null hypothesis that the quadratic growth model is adequate becomes

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{W}_2|} = \frac{\left|\begin{bmatrix} 2726.282 & 2660.749 & 2369.308 & 2335.912 \\ 2660.749 & 2756.009 & 2343.514 & 2327.961 \\ 2369.308 & 2343.514 & 2301.714 & 2098.544 \\ 2335.912 & 2327.961 & 2098.544 & 2277.452 \end{bmatrix}\right|}{\left|\begin{bmatrix} 2781.017 & 2698.589 & 2363.228 & 2362.253 \\ 2698.589 & 2832.430 & 2331.235 & 2381.160 \\ 2363.228 & 2331.235 & 2303.687 & 2089.996 \\ 2362.253 & 2381.160 & 2089.996 & 2314.485 \end{bmatrix}\right|} = .7627$$

Since, with $\alpha = .01$,

$$-\left(N - \frac{1}{2}(p - q + g)\right) \ln \Lambda^* = -\left(31 - \frac{1}{2}(4 - 2 + 2)\right) \ln .7627$$

$$= 7.86 < \chi^2_{(4-2-1)2}(.01) = 9.21$$

we fail to reject the adequacy of the quadratic fit at $\alpha = .01$. Since the p-value is less than .05 there is, however, some evidence that the quadratic does not fit well.

We could, without restricting to quadratic growth, test for parallel and coincident calcium loss using profile analysis. ■

The Potthoff and Roy growth curve model holds for more general designs than one-way MANOVA. However, the $\hat{\boldsymbol{\beta}}_\ell$ are no longer given by (6-78) and the expression for its covariance matrix becomes more complicated than (6-79). We refer the reader to [14] for more examples and further tests.

There are many other modifications to the model treated here. They include the following:

(a) Dropping the restriction to polynomial growth. Use nonlinear parametric models or even nonparametric splines.

(b) Restricting the covariance matrix to a special form such as equally correlated responses on the same individual.

(c) Observing more than one response variable, over time, on the same individual. This results in a multivariate version of the growth curve model.

6.10 Perspectives and a Strategy for Analyzing Multivariate Models

We emphasize that, with several characteristics, it is important to control the overall probability of making any incorrect decision. This is particularly important when testing for the equality of two or more treatments as the examples in this chapter indicate. A single multivariate test, with its associated single p-value, is preferable to performing a large number of univariate tests. The outcome tells us whether or not it is worthwhile to look closer on a variable by variable and group by group analysis.

A single multivariate test is recommended over, say, p univariate tests because, as the next example demonstrates, univariate tests ignore important information and can give misleading results.

Example 6.16 (Comparing multivariate and univariate tests for the differences in means) Suppose we collect measurements on two variables X_1 and X_2 for ten randomly selected experimental units from each of two groups. The hypothetical data are noted here and displayed as scatter plots and marginal dot diagrams in Figure 6.6 on page 334.

x_1	x_2	Group
5.0	3.0	1
4.5	3.2	1
6.0	3.5	1
6.0	4.6	1
6.2	5.6	1
6.9	5.2	1
6.8	6.0	1
5.3	5.5	1
6.6	7.3	1
7.3	6.5	1
4.6	4.9	2
4.9	5.9	2
4.0	4.1	2
3.8	5.4	2
6.2	6.1	2
5.0	7.0	2
5.3	4.7	2
7.1	6.6	2
5.8	7.8	2
6.8	8.0	2

It is clear from the horizontal marginal dot diagram that there is considerable overlap in the x_1 values for the two groups. Similarly, the vertical marginal dot diagram shows there is considerable overlap in the x_2 values for the two groups. The scatter plots suggest that there is fairly strong positive correlation between the two variables for each group, and that, although there is some overlap, the group 1 measurements are generally to the southeast of the group 2 measurements.

Let $\boldsymbol{\mu}_1' = [\mu_{11}, \mu_{12}]$ be the population mean vector for the first group, and let $\boldsymbol{\mu}_2' = [\mu_{21}, \mu_{22}]$ be the population mean vector for the second group. Using the x_1 observations, a univariate analysis of variance gives $F = 2.46$ with $\nu_1 = 1$ and $\nu_2 = 18$ degrees of freedom. Consequently, we cannot reject $H_0: \mu_{11} = \mu_{21}$ at any reasonable significance level ($F_{1,18}(.10) = 3.01$). Using the x_2 observations, a univariate analysis of variance gives $F = 2.68$ with $\nu_1 = 1$ and $\nu_2 = 18$ degrees of freedom. Again, we cannot reject $H_0: \mu_{12} = \mu_{22}$ at any reasonable significance level.

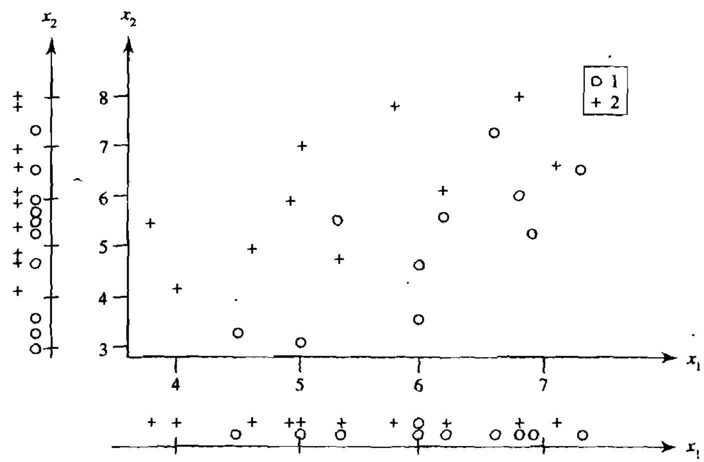


Figure 6.6 Scatter plots and marginal dot diagrams for the data from two groups.

The univariate tests suggest there is no difference between the component means for the two groups, and hence we cannot discredit $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$.

On the other hand, if we use Hotelling's T^2 to test for the equality of the mean vectors, we find

$$T^2 = 17.29 > c^2 = \frac{(18)(2)}{17} F_{2,17}(.01) = 2.118 \times 6.11 = 12.94$$

and we reject $H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ at the 1% level. The multivariate test takes into account the positive correlation between the two measurements for each group—information that is unfortunately ignored by the univariate tests. This T^2-test is equivalent to the MANOVA test (6-42). ■

Example 6.17 (Data on lizards that require a bivariate test to establish a difference in means) A zoologist collected lizards in the southwestern United States. Among other variables, he measured mass (in grams) and the snout-vent length (in millimeters). Because the tails sometimes break off in the wild, the snout-vent length is a more representative measure of length. The data for the lizards from two genera, Cnemidophorus (C) and Sceloporus (S), collected in 1997 and 1999 are given in Table 6.7. Notice that there are $n_1 = 20$ measurements for C lizards and $n_2 = 40$ measurements for S lizards.

After taking natural logarithms, the summary statistics are

$$\text{C}: n_1 = 20 \qquad \bar{\mathbf{x}}_1 = \begin{bmatrix} 2.240 \\ 4.394 \end{bmatrix} \qquad \mathbf{S}_1 = \begin{bmatrix} 0.35305 & 0.09417 \\ 0.09417 & 0.02595 \end{bmatrix}$$

$$\text{S}: n_2 = 40 \qquad \bar{\mathbf{x}}_2 = \begin{bmatrix} 2.368 \\ 4.308 \end{bmatrix} \qquad \mathbf{S}_2 = \begin{bmatrix} 0.50684 & 0.14539 \\ 0.14539 & 0.04255 \end{bmatrix}$$

Table 6.7 Lizard Data for Two Genera

C		S		S	
Mass	SVL	Mass	SVL	Mass	SVL
7.513	74.0	13.911	77.0	14.666	80.0
5.032	69.5	5.236	62.0	4.790	62.0
5.867	72.0	37.331	108.0	5.020	61.5
11.088	80.0	41.781	115.0	5.220	62.0
2.419	56.0	31.995	106.0	5.690	64.0
13.610	94.0	3.962	56.0	6.763	63.0
18.247	95.5	4.367	60.5	9.977	71.0
16.832	99.5	3.048	52.0	8.831	69.5
15.910	97.0	4.838	60.0	9.493	67.5
17.035	90.5	6.525	64.0	7.811	66.0
16.526	91.0	22.610	96.0	6.685	64.5
4.530	67.0	13.342	79.5	11.980	79.0
7.230	75.0	4.109	55.5	16.520	84.0
5.200	69.5	12.369	75.0	13.630	81.0
13.450	91.5	7.120	64.5	13.700	82.5
14.080	91.0	21.077	87.5	10.350	74.0
14.665	90.0	42.989	109.0	7.900	68.5
6.092	73.0	27.201	96.0	9.103	70.0
5.264	69.5	38.901	111.0	13.216	77.5
16.902	94.0	19.747	84.5	9.787	70.0

SVL = snout-vent length.
Source: Data courtesy of Kevin E. Bonine.

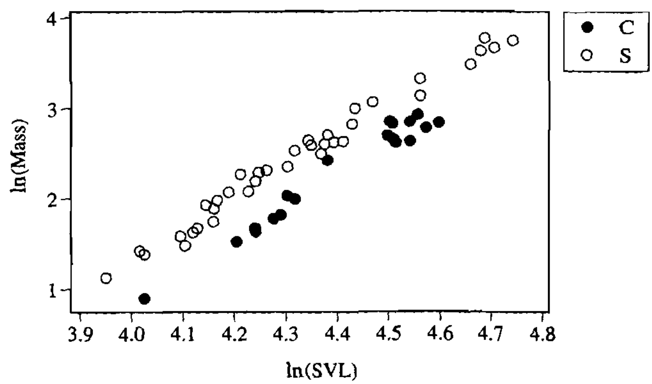


Figure 6.7 Scatter plot of ln(Mass) versus ln(SVL) for the lizard data in Table 6.7.

A plot of mass (Mass) versus snout-vent length (SVL), after taking natural logarithms, is shown in Figure 6.7. The large sample individual 95% confidence intervals for the difference in ln(Mass) means and the difference in ln(SVL) means both cover 0.

$$\text{ln(Mass)}: \quad \mu_{11} - \mu_{21}: \quad (-0.476, 0.220)$$
$$\text{ln(SVL)}: \quad \mu_{12} - \mu_{22}: \quad (-0.011, 0.183)$$

The corresponding univariate Student's t-test statistics for testing for no difference in the individual means have p-values of .46 and .08, respectively. Clearly, from a univariate perspective, we cannot detect a difference in mass means or a difference in snout-vent length means for the two genera of lizards.

However, consistent with the scatter diagram in Figure 6.7, a bivariate analysis strongly supports a difference in size between the two groups of lizards. Using Result 6.4 (also see Example 6.5), the T^2-statistic has an approximate χ_2^2 distribution. For this example, $T^2 = 225.4$ with a p-value less than .0001. A multivariate method is essential in this case. ■

Examples 6.16 and 6.17 demonstrate the efficacy of a multivariate test relative to its univariate counterparts. We encountered exactly this situation with the effluent data in Example 6.1.

In the context of random samples from several populations (recall the one-way MANOVA in Section 6.4), multivariate tests are based on the matrices

$$\mathbf{W} = \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)' \quad \text{and} \quad \mathbf{B} = \sum_{\ell=1}^{g} n_\ell (\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})'$$

Throughout this chapter, we have used

$$\text{Wilks' lambda statistic } \Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|}$$

which is equivalent to the likelihood ratio test. Three other multivariate test statistics are regularly included in the output of statistical packages.

$$\text{Lawley–Hotelling trace} = \operatorname{tr}[\mathbf{B}\mathbf{W}^{-1}]$$

$$\text{Pillai trace} = \operatorname{tr}[\mathbf{B}(\mathbf{B} + \mathbf{W})^{-1}]$$

$$\text{Roy's largest root} = \text{maximum eigenvalue of } \mathbf{W}(\mathbf{B} + \mathbf{W})^{-1}$$

All four of these tests appear to be nearly equivalent for extremely large samples. For moderate sample sizes, all comparisons are based on what is necessarily a limited number of cases studied by simulation. From the simulations reported to date, the first three tests have similar power, while the last, Roy's test, behaves differently. Its power is best only when there is a single nonzero eigenvalue and, at the same time, the power is large. This may approximate situations where a large difference exists in just one characteristic and it is between one group and all of the others. There is also some suggestion that Pillai's trace is slightly more robust against nonnormality. However, we suggest trying transformations on the original data when the residuals are nonnormal.

All four statistics apply in the two-way setting and in even more complicated MANOVA. More discussion is given in terms of the multivariate regression model in Chapter 7.

When, and only when, the multivariate tests signals a difference, or departure from the null hypothesis, do we probe deeper. We recommend calculating the Bonferonni intervals for all pairs of groups and all characteristics. The simultaneous confidence statements determined from the shadows of the confidence ellipse are, typically, too large. The one-at-a-time intervals may be suggestive of differences that merit further study but, with the current data, cannot be taken as conclusive evidence for the existence of differences. We summarize the procedure developed in this chapter for comparing treatments. The first step is to check the data for outliers using visual displays and other calculations.

A Strategy for the Multivariate Comparison of Treatments

1. *Try to identify outliers.* Check the data group by group for outliers. Also check the collection of residual vectors from any fitted model for outliers. Be aware of any outliers so calculations can be performed with and without them.
2. *Perform a multivariate test of hypothesis.* Our choice is the likelihood ratio test, which is equivalent to Wilks' lambda test.
3. *Calculate the Bonferroni simultaneous confidence intervals.* If the multivariate test reveals a difference, then proceed to calculate the Bonferroni confidence intervals for all pairs of groups or treatments, and all characteristics. If no differences are significant, try looking at Bonferroni intervals for the larger set of responses that includes the differences and sums of pairs of responses.

We must issue one caution concerning the proposed strategy. It may be the case that differences would appear in only one of the many characteristics and, further, the differences hold for only a few treatment combinations. Then, these few active differences may become lost among all the inactive ones. That is, the overall test may not show significance whereas a univariate test restricted to the specific active variable would detect the difference. The best preventative is a good experimental design. To design an effective experiment when one specific variable is expected to produce differences, do not include too many other variables that are not expected to show differences among the treatments.

Exercises

6.1. Construct and sketch a joint 95% confidence region for the mean difference vector $\boldsymbol{\delta}$ using the effluent data and results in Example 6.1. Note that the point $\boldsymbol{\delta} = \mathbf{0}$ falls outside the 95% contour. Is this result consistent with the test of H_0: $\boldsymbol{\delta} = \mathbf{0}$ considered in Example 6.1? Explain.

6.2. Using the information in Example 6.1. construct the 95% Bonferroni simultaneous intervals for the components of the mean difference vector $\boldsymbol{\delta}$. Compare the lengths of these intervals with those of the simultaneous intervals constructed in the example.

6.3. The data corresponding to sample 8 in Table 6.1 seem unusually large. Remove sample 8. Construct a joint 95% confidence region for the mean difference vector $\boldsymbol{\delta}$ and the 95% Bonferroni simultaneous intervals for the components of the mean difference vector. Are the results consistent with a test of H_0: $\boldsymbol{\delta} = \mathbf{0}$? Discuss. Does the "outlier" make a difference in the analysis of these data?

6.4. Refer to Example 6.1.

(a) Redo the analysis in Example 6.1 after transforming the pairs of observations to $\ln(\text{BOD})$ and $\ln(\text{SS})$.

(b) Construct the 95% Bonferroni simultaneous intervals for the components of the mean vector $\boldsymbol{\delta}$ of transformed variables.

(c) Discuss any possible violation of the assumption of a bivariate normal distribution for the difference vectors of transformed observations.

6.5. A researcher considered three indices measuring the severity of heart attacks. The values of these indices for $n = 40$ heart-attack patients arriving at a hospital emergency room produced the summary statistics

$$\bar{\mathbf{x}} = \begin{bmatrix} 46.1 \\ 57.3 \\ 50.4 \end{bmatrix} \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} 101.3 & 63.0 & 71.0 \\ 63.0 & 80.2 & 55.6 \\ 71.0 & 55.6 & 97.4 \end{bmatrix}$$

(a) All three indices are evaluated for each patient. Test for the equality of mean indices using (6-16) with $\alpha = .05$.

(b) Judge the differences in pairs of mean indices using 95% simultaneous confidence intervals. [See (6-18).]

6.6. Use the data for treatments 2 and 3 in Exercise 6.8.

(a) Calculate $\mathbf{S}_{\text{pooled}}$.

(b) Test $H_0: \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3 = \mathbf{0}$ employing a two-sample approach with $\alpha = .01$.

(c) Construct 99% simultaneous confidence intervals for the differences $\mu_{2i} - \mu_{3i}$, $i = 1, 2$.

6.7. Using the summary statistics for the electricity-demand data given in Example 6.4, compute T^2 and test the hypothesis $H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$, assuming that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. Set $\alpha = .05$. Also, determine the linear combination of mean components most responsible for the rejection of H_0.

6.8. Observations on two responses are collected for three treatments. The observation vectors $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ are

$$\text{Treatment 1:} \quad \begin{bmatrix} 6 \\ 7 \end{bmatrix}, \begin{bmatrix} 5 \\ 9 \end{bmatrix}, \begin{bmatrix} 8 \\ 6 \end{bmatrix}, \begin{bmatrix} 4 \\ 9 \end{bmatrix}, \begin{bmatrix} 7 \\ 9 \end{bmatrix}$$

$$\text{Treatment 2:} \quad \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\text{Treatment 3:} \quad \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

(a) Break up the observations into mean, treatment, and residual components, as in (6-39). Construct the corresponding arrays for each variable. (See Example 6.9.)

(b) Using the information in Part a, construct the one-way MANOVA table.

(c) Evaluate Wilks' lambda, Λ^*, and use Table 6.3 to test for treatment effects. Set $\alpha = .01$. Repeat the test using the chi-square approximation with Bartlett's correction. [See (6-43).] Compare the conclusions.

6.9. Using the contrast matrix $\mathbf{C}$ in (6-13), verify the relationships $\mathbf{d}_j = \mathbf{C}\mathbf{x}_j$, $\bar{\mathbf{d}} = \mathbf{C}\bar{\mathbf{x}}$, and $\mathbf{S}_d = \mathbf{C}\mathbf{S}\mathbf{C}'$ in (6-14).

6.10. Consider the univariate one-way decomposition of the observation $x_{\ell j}$ given by (6-34). Show that the mean vector $\bar{x}\,\mathbf{1}$ is always perpendicular to the treatment effect vector $(\bar{x}_1 - \bar{x})\mathbf{u}_1 + (\bar{x}_2 - \bar{x})\mathbf{u}_2 + \cdots + (\bar{x}_g - \bar{x})\mathbf{u}_g$ where

$$\mathbf{u}_1 = \begin{bmatrix} \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\} n_1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\} n_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{u}_g = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\} n_g \end{bmatrix}$$

6.11. A likelihood argument provides additional support for pooling the two independent sample covariance matrices to estimate a common covariance matrix in the case of two normal populations. Give the likelihood function, $L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, for two independent samples of sizes n_1 and n_2 from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ populations, respectively. Show that this likelihood is maximized by the choices $\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{x}}_1$, $\hat{\boldsymbol{\mu}}_2 = \bar{\mathbf{x}}_2$ and

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n_1 + n_2}[(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2] = \left(\frac{n_1 + n_2 - 2}{n_1 + n_2}\right)\mathbf{S}_{\text{pooled}}$$

Hint: Use (4-16) and the maximization Result 4.10.

6.12. *(Test for linear profiles, given that the profiles are parallel.)* Let $\boldsymbol{\mu}_1' = [\mu_{11}, \mu_{12}, \ldots, \mu_{1p}]$ and $\boldsymbol{\mu}_2' = [\mu_{21}, \mu_{22}, \ldots, \mu_{2p}]$ be the mean responses to p treatments for populations 1 and 2, respectively. Assume that the profiles given by the two mean vectors are parallel.

(a) Show that the hypothesis that the profiles are linear can be written as H_0: $(\mu_{1i} + \mu_{2i}) - (\mu_{1i-1} + \mu_{2i-1}) = (\mu_{1i-1} + \mu_{2i-1}) - (\mu_{1i-2} + \mu_{2i-2})$, $i = 3, \ldots, p$ or as H_0: $\mathbf{C}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = \mathbf{0}$, where the $(p - 2) \times p$ matrix

$$\mathbf{C} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix}$$

(b) Following an argument similar to the one leading to (6-73), we reject H_0: $\mathbf{C}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = \mathbf{0}$ at level α if

$$T^2 = (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)'\mathbf{C}'\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{C}\mathbf{S}_{\text{pooled}}\mathbf{C}'\right]^{-1}\mathbf{C}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) > c^2$$

where

$$c^2 = \frac{(n_1 + n_2 - 2)(p - 2)}{n_1 + n_2 - p + 1} F_{p-2, n_1+n_2-p+1}(\alpha)$$

Let $n_1 = 30$, $n_2 = 30$, $\bar{\mathbf{x}}_1' = [6.4, 6.8, 7.3, 7.0]$, $\bar{\mathbf{x}}_2' = [4.3, 4.9, 5.3, 5.1]$, and

$$\mathbf{S}_{\text{pooled}} = \begin{bmatrix} .61 & .26 & .07 & .16 \\ .26 & .64 & .17 & .14 \\ .07 & .17 & .81 & .03 \\ .16 & .14 & .03 & .31 \end{bmatrix}$$

Test for linear profiles, assuming that the profiles are parallel. Use $\alpha = .05$.

6.13. *(Two-way MANOVA without replications.)* Consider the observations on two responses, x_1 and x_2, displayed in the form of the following two-way table (note that there is a *single* observation vector at each combination of factor levels):

		Factor 2 Level 1	Factor 2 Level 2	Factor 2 Level 3	Factor 2 Level 4
	Level 1	$\begin{bmatrix} 6 \\ 8 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 6 \end{bmatrix}$	$\begin{bmatrix} 8 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 6 \end{bmatrix}$
Factor 1	Level 2	$\begin{bmatrix} 3 \\ 8 \end{bmatrix}$	$\begin{bmatrix} -3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 3 \end{bmatrix}$	$\begin{bmatrix} -4 \\ 3 \end{bmatrix}$
	Level 3	$\begin{bmatrix} -3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} -4 \\ -5 \end{bmatrix}$	$\begin{bmatrix} 3 \\ -3 \end{bmatrix}$	$\begin{bmatrix} -4 \\ -6 \end{bmatrix}$

With no replications, the two-way MANOVA model is

$$\mathbf{X}_{\ell k} = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell + \boldsymbol{\beta}_k + \mathbf{e}_{\ell k}; \qquad \sum_{\ell=1}^{g} \boldsymbol{\tau}_\ell = \sum_{k=1}^{b} \boldsymbol{\beta}_k = \mathbf{0}$$

where the $\mathbf{e}_{\ell k}$ are independent $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ random vectors.

(a) Decompose the observations for each of the two variables as

$$x_{\ell k} = \bar{x} + (\bar{x}_{\ell\cdot} - \bar{x}) + (\bar{x}_{\cdot k} - \bar{x}) + (x_{\ell k} - \bar{x}_{\ell\cdot} - \bar{x}_{\cdot k} + \bar{x})$$

similar to the arrays in Example 6.9. *For each response*, this decomposition will result in several 3×4 matrices. Here $\bar{x}$ is the overall average, $\bar{x}_{\ell\cdot}$ is the average for the ℓth level of factor 1, and $\bar{x}_{\cdot k}$ is the average for the kth level of factor 2.

(b) Regard the rows of the matrices in Part a as strung out in a single "long" vector, and compute the sums of squares

$$SS_{\text{tot}} = SS_{\text{mean}} + SS_{\text{fac1}} + SS_{\text{fac2}} + SS_{\text{res}}$$

and sums of cross products

$$SCP_{\text{tot}} = SCP_{\text{mean}} + SCP_{\text{fac1}} + SCP_{\text{fac2}} + SCP_{\text{res}}$$

Consequently, obtain the matrices $\mathbf{SSP}_{\text{cor}}$, $\mathbf{SSP}_{\text{fac1}}$, $\mathbf{SSP}_{\text{fac2}}$, and $\mathbf{SSP}_{\text{res}}$ with degrees of freedom $gb - 1$, $g - 1$, $b - 1$, and $(g - 1)(b - 1)$, respectively.

(c) Summarize the calculations in Part b in a MANOVA table.

Hint: This MANOVA table is consistent with the two-way MANOVA table for comparing factors and their interactions where $n = 1$. Note that, with $n = 1$, $\mathbf{SSP}_{\text{res}}$ in the general two-way MANOVA table is a zero matrix with zero degrees of freedom. The matrix of interaction sum of squares and cross products now becomes the *residual* sum of squares and cross products matrix.

(d) Given the summary in Part c, test for factor 1 and factor 2 main effects at the $\alpha = .05$ level.

Hint: Use the results in (6-67) and (6-69) with $gb(n - 1)$ replaced by $(g - 1)(b - 1)$.
Note: The tests require that $p \le (g - 1)(b - 1)$ so that $\mathbf{SSP}_{\text{res}}$ will be positive definite (with probability 1).

6.14. A *replicate* of the experiment in Exercise 6.13 yields the following data:

		Factor 2			
		Level 1	Level 2	Level 3	Level 4
	Level 1	$\begin{bmatrix} 14 \\ 8 \end{bmatrix}$	$\begin{bmatrix} 6 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 8 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 16 \\ -4 \end{bmatrix}$
Factor 1	Level 2	$\begin{bmatrix} 1 \\ 6 \end{bmatrix}$	$\begin{bmatrix} 5 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 15 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 7 \end{bmatrix}$
	Level 3	$\begin{bmatrix} 3 \\ -2 \end{bmatrix}$	$\begin{bmatrix} -2 \\ 7 \end{bmatrix}$	$\begin{bmatrix} -11 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -6 \\ 6 \end{bmatrix}$

(a) Use these data to decompose each of the two measurements in the observation vector as

$$x_{\ell k} = \bar{x} + (\bar{x}_{\ell\cdot} - \bar{x}) + (\bar{x}_{\cdot k} - \bar{x}) + (x_{\ell k} - \bar{x}_{\ell\cdot} - \bar{x}_{\cdot k} + \bar{x})$$

where $\bar{x}$ is the overall average, $\bar{x}_{\ell\cdot}$ is the average for the ℓth level of factor 1, and $\bar{x}_{\cdot k}$ is the average for the kth level of factor 2. Form the corresponding arrays for each of the two responses.

(b) Combine the preceding data with the data in Exercise 6.13 and carry out the necessary calculations to complete the general two-way MANOVA table.

(c) Given the results in Part b, test for interactions, and if the interactions do not exist, test for factor 1 and factor 2 main effects. Use the likelihood ratio test with $\alpha = .05$.

(d) If main effects, but no interactions, exist, examine the nature of the main effects by constructing Bonferroni simultaneous 95% confidence intervals for differences of the components of the factor effect parameters.

6.15. Refer to Example 6.13.

(a) Carry out approximate chi-square (likelihood ratio) tests for the factor 1 and factor 2 effects. Set $\alpha = .05$. Compare these results with the results for the exact F-tests given in the example. Explain any differences.

(b) Using (6-70), construct simultaneous 95% confidence intervals for differences in the factor 1 effect parameters for *pairs* of the three responses. Interpret these intervals. Repeat these calculations for factor 2 effect parameters.

The following exercises may require the use of a computer.

6.16. Four measures of the response *stiffness* on each of 30 boards are listed in Table 4.3 (see Example 4.14). The measures, on a given board, are repeated in the sense that they were made one after another. Assuming that the measures of stiffness arise from four treatments, test for the equality of treatments in a *repeated measures design* context. Set $\alpha = .05$. Construct a 95% (simultaneous) confidence interval for a contrast in the mean levels representing a comparison of the dynamic measurements with the static measurements.

6.17. The data in Table 6.8 were collected to test two psychological models of numerical cognition. Does the processing of numbers depend on the way the numbers are presented (words, Arabic digits)? Thirty-two subjects were required to make a series of

Table 6.8 Number Parity Data (Median Times in Milliseconds)

WordDiff (x_1)	WordSame (x_2)	ArabicDiff (x_3)	ArabicSame (x_4)
869.0	860.5	691.0	601.0
995.0	875.0	678.0	659.0
1056.0	930.5	833.0	826.0
1126.0	954.0	888.0	728.0
1044.0	909.0	865.0	839.0
925.0	856.5	1059.5	797.0
1172.5	896.5	926.0	766.0
1408.5	1311.0	854.0	986.0
1028.0	887.0	915.0	735.0
1011.0	863.0	761.0	657.0
726.0	674.0	663.0	583.0
982.0	894.0	831.0	640.0
1225.0	1179.0	1037.0	905.5
731.0	662.0	662.5	624.0
975.5	872.5	814.0	735.0
1130.5	811.0	843.0	657.0
945.0	909.0	867.5	754.0
747.0	752.5	777.0	687.5
656.5	659.5	572.0	539.0
919.0	833.0	752.0	611.0
751.0	744.0	683.0	553.0
774.0	735.0	671.0	612.0
941.0	931.0	901.5	700.0
751.0	785.0	789.0	735.0
767.0	737.5	724.0	639.0
813.5	750.5	711.0	625.0
1289.5	1140.0	904.5	784.5
1096.5	1009.0	1076.0	983.0
1083.0	958.0	918.0	746.5
1114.0	1046.0	1081.0	796.0
708.0	669.0	657.0	572.5
1201.0	925.0	1004.5	673.5

Source: Data courtesy of J. Carr.

quick numerical judgments about two numbers presented as either two number words ("two," "four") or two single Arabic digits ("2," "4"). The subjects were asked to respond "same" if the two numbers had the same numerical parity (both even or both odd) and "different" if the two numbers had a different parity (one even, one odd). Half of the subjects were assigned a block of Arabic digit trials, followed by a block of number word trials, and half of the subjects received the blocks of trials in the reverse order. Within each block, the order of "same" and "different" parity trials was randomized for each subject. For each of the four combinations of parity and format, the median reaction times for correct responses were recorded for each subject. Here

X_1 = median reaction time for word format–different parity combination
X_2 = median reaction time for word format–same parity combination
X_3 = median reaction time for Arabic format–different parity combination
X_4 = median reaction time for Arabic format–same parity combination

(a) Test for treatment effects using a *repeated measures design*. Set $\alpha = .05$.

(b) Construct 95% (simultaneous) confidence intervals for the contrasts representing the number format effect, the parity type effect and the interaction effect. Interpret the resulting intervals.

(c) The absence of interaction supports the M model of numerical cognition, while the presence of interaction supports the C and C model of numerical cognition. Which model is supported in this experiment?

(d) For each subject, construct three difference scores corresponding to the number format contrast, the parity type contrast, and the interaction contrast. Is a multivariate normal distribution a reasonable population model for these data? Explain.

6.18. Jolicoeur and Mosimann [12] studied the relationship of size and shape for painted turtles. Table 6.9 contains their measurements on the carapaces of 24 female and 24 male turtles.

(a) Test for equality of the two population mean vectors using $\alpha = .05$.

(b) If the hypothesis in Part a is rejected, find the linear combination of mean components most responsible for rejecting H_0.

(c) Find simultaneous confidence intervals for the component mean differences. Compare with the Bonferroni intervals.

Hint: You may wish to consider logarithmic transformations of the observations.

6.19. In the first phase of a study of the cost of transporting milk from farms to dairy plants, a survey was taken of firms engaged in milk transportation. Cost data on X_1 = fuel, X_2 = repair, and X_3 = capital, all measured on a per-mile basis, are presented in Table 6.10 on page 345 for $n_1 = 36$ gasoline and $n_2 = 23$ diesel trucks.

(a) Test for differences in the mean cost vectors. Set $\alpha = .01$.

(b) If the hypothesis of equal cost vectors is rejected in Part a, find the linear combination of mean components most responsible for the rejection.

(c) Construct 99% simultaneous confidence intervals for the pairs of mean components. Which costs, if any, appear to be quite different?

(d) Comment on the validity of the assumptions used in your analysis. Note in particular that observations 9 and 21 for gasoline trucks have been identified as multivariate outliers. (See Exercise 5.22 and [2].) Repeat Part a with these observations deleted. Comment on the results.

Table 6.9 Carapace Measurements (in Millimeters) for Painted Turtles

Female			Male		
Length (x_1)	Width (x_2)	Height (x_3)	Length (x_1)	Width (x_2)	Height (x_3)
98	81	38	93	74	37
103	84	38	94	78	35
103	86	42	96	80	35
105	86	42	101	84	39
109	88	44	102	85	38
123	92	50	103	81	37
123	95	46	104	83	39
133	99	51	106	83	39
133	102	51	107	82	38
133	102	51	112	89	40
134	100	48	113	88	40
136	102	49	114	86	40
138	98	51	116	90	43
138	99	51	117	90	41
141	105	53	117	91	41
147	108	57	119	93	41
149	107	55	120	89	40
153	107	56	120	93	44
155	115	63	121	95	42
155	117	60	125	93	45
158	115	62	127	96	45
159	118	63	128	95	45
162	124	61	131	95	46
177	132	67	135	106	47

6.20. The tail lengths in millimeters (x_1) and wing lengths in millimeters (x_2) for 45 *male* hook-billed kites are given in Table 6.11 on page 346. Similar measurements for female hook-billed kites were given in Table 5.12.

(a) Plot the male hook-billed kite data as a scatter diagram, and (visually) check for outliers. (Note, in particular, observation 31 with $x_1 = 284$.)

(b) Test for equality of mean vectors for the populations of male and female hook-billed kites. Set $\alpha = .05$. If H_0: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ is rejected, find the linear combination most responsible for the rejection of H_0. (You may want to eliminate any outliers found in Part a for the male hook-billed kite data before conducting this test. Alternatively, you may want to interpret $x_1 = 284$ for observation 31 as a misprint and conduct the test with $x_1 = 184$ for this observation. Does it make any difference in this case how observation 31 for the male hook-billed kite data is treated?)

(c) Determine the 95% confidence region for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ and 95% simultaneous confidence intervals for the components of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.

(d) Are male or female birds generally larger?

Table 6.10 Milk Transportation-Cost Data

Gasoline trucks			Diesel trucks		
x_1	x_2	x_3	x_1	x_2	x_3
16.44	12.43	11.23	8.50	12.26	9.11
7.19	2.70	3.92	7.42	5.13	17.15
9.92	1.35	9.75	10.28	3.32	11.23
4.24	5.78	7.78	10.16	14.72	5.99
11.20	5.05	10.67	12.79	4.17	29.28
14.25	5.78	9.88	9.60	12.72	11.00
13.50	10.98	10.60	6.47	8.89	19.00
13.32	14.27	9.45	11.35	9.95	14.53
29.11	15.09	3.28	9.15	2.94	13.68
12.68	7.61	10.23	9.70	5.06	20.84
7.51	5.80	8.13	9.77	17.86	35.18
9.90	3.63	9.13	11.61	11.75	17.00
10.25	5.07	10.17	9.09	13.25	20.66
11.11	6.15	7.61	8.53	10.14	17.45
12.17	14.26	14.39	8.29	6.22	16.38
10.24	2.59	6.09	15.90	12.90	19.09
10.18	6.05	12.14	11.94	5.69	14.77
8.88	2.70	12.23	9.54	16.77	22.66
12.34	7.73	11.68	10.43	17.65	10.66
8.51	14.02	12.01	10.87	21.52	28.47
26.16	17.44	16.89	7.13	13.22	19.44
12.95	8.24	7.18	11.88	12.18	21.20
16.93	13.37	17.59	12.03	9.22	23.09
14.70	10.78	14.58			
10.32	5.16	17.00			
8.98	4.49	4.26			
9.70	11.59	6.83			
12.72	8.63	5.59			
9.49	2.16	6.23			
8.22	7.95	6.72			
13.70	11.22	4.91			
8.21	9.85	8.17			
15.86	11.42	13.06			
9.18	9.18	9.49			
12.49	4.67	11.94			
17.32	6.86	4.44			

Source: Data courtesy of M. Keaton.

6.21. Using Moody's bond ratings, samples of 20 Aa (middle-high quality) corporate bonds and 20 Baa (top-medium quality) corporate bonds were selected. For each of the corresponding companies, the ratios

X_1 = current ratio (a measure of short-term liquidity)

X_2 = long-term interest rate (a measure of interest coverage)

X_3 = debt-to-equity ratio (a measure of financial risk or leverage)

X_4 = rate of return on equity (a measure of profitability)

Table 6.11 Male Hook-Billed Kite Data

x_1 (Tail length)	x_2 (Wing length)	x_1 (Tail length)	x_2 (Wing length)	x_1 (Tail length)	x_2 (Wing length)
180	278	185	282	284	277
186	277	195	285	176	281
206	308	183	276	185	287
184	290	202	308	191	295
177	273	177	254	177	267
177	284	177	268	197	310
176	267	170	260	199	299
200	281	186	274	190	273
191	287	177	272	180	278
193	271	178	266	189	280
212	302	192	281	194	290
181	254	204	276	186	287
195	297	191	290	191	286
187	281	178	265	187	288
190	284	177	275	186	275

Source: Data courtesy of S. Temple.

were recorded. The summary statistics are as follows:

Aa bond companies: $n_1 = 20, \bar{\mathbf{x}}_1' = [2.287, 12.600, .347, 14.830]$, and

$$\mathbf{S}_1 = \begin{bmatrix} .459 & .254 & -.026 & -.244 \\ .254 & 27.465 & -.589 & -.267 \\ -.026 & -.589 & .030 & .102 \\ -.244 & -.267 & .102 & 6.854 \end{bmatrix}$$

Baa bond companies: $n_2 = 20, \bar{\mathbf{x}}_2' = [2.404, 7.155, .524, 12.840]$,

$$\mathbf{S}_2 = \begin{bmatrix} .944 & -.089 & .002 & -.719 \\ -.089 & 16.432 & -.400 & 19.044 \\ .002 & -.400 & .024 & -.094 \\ -.719 & 19.044 & -.094 & 61.854 \end{bmatrix}$$

and

$$\mathbf{S}_{\text{pooled}} = \begin{bmatrix} .701 & .083 & -.012 & -.481 \\ .083 & 21.949 & -.494 & 9.388 \\ -.012 & -.494 & .027 & .004 \\ -.481 & 9.388 & .004 & 34.354 \end{bmatrix}$$

(a) Does pooling appear reasonable here? Comment on the pooling procedure in this case.

(b) Are the financial characteristics of firms with Aa bonds different from those with Baa bonds? Using the pooled covariance matrix, test for the equality of mean vectors. Set $\alpha = .05$.

(c) Calculate the linear combinations of mean components most responsible for rejecting $H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ in Part b.

(d) Bond rating companies are interested in a company's ability to satisfy its outstanding debt obligations as they mature. Does it appear as if one or more of the foregoing financial ratios might be useful in helping to classify a bond as "high" or "medium" quality? Explain.

(e) Repeat part (b) assuming normal populations with unequal covariance matices (see (6-27), (6-28) and (6-29)). Does your conclusion change?

6.22. Researchers interested in assessing pulmonary function in nonpathological populations asked subjects to run on a treadmill until exhaustion. Samples of air were collected at definite intervals and the gas contents analyzed. The results on 4 measures of oxygen consumption for 25 males and 25 females are given in Table 6.12 on page 348. The variables were

$$\begin{aligned} X_1 &= \text{resting volume } O_2 \text{ (L/min)} \\ X_2 &= \text{resting volume } O_2 \text{ (mL/kg/min)} \\ X_3 &= \text{maximum volume } O_2 \text{ (L/min)} \\ X_4 &= \text{maximum volume } O_2 \text{ (mL/kg/min)} \end{aligned}$$

(a) Look for gender differences by testing for equality of group means. Use $\alpha = .05$. If you reject $H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$, find the linear combination most responsible.

(b) Construct the 95% simultaneous confidence intervals for each $\mu_{1i} - \mu_{2i}$, $i = 1, 2, 3, 4$. Compare with the corresponding Bonferroni intervals.

(c) The data in Table 6.12 were collected from graduate-student volunteers, and thus they do not represent a random sample. Comment on the possible implications of this information.

6.23. Construct a one-way MANOVA using the width measurements from the iris data in Table 11.5. Construct 95% simultaneous confidence intervals for differences in mean components for the two responses for each pair of populations. Comment on the validity of the assumption that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3$.

6.24. Researchers have suggested that a change in skull size over time is evidence of the interbreeding of a resident population with immigrant populations. Four measurements were made of male Egyptian skulls for three different time periods: period 1 is 4000 B.C., period 2 is 3300 B.C., and period 3 is 1850 B.C. The data are shown in Table 6.13 on page 349 (see the skull data on the website www.prenhall.com/statistics). The measured variables are

$$\begin{aligned} X_1 &= \text{maximum breadth of skull (mm)} \\ X_2 &= \text{basibregmatic height of skull (mm)} \\ X_3 &= \text{basialveolar length of skull (mm)} \\ X_4 &= \text{nasal height of skull (mm)} \end{aligned}$$

Construct a one-way MANOVA of the Egyptian skull data. Use $\alpha = .05$. Construct 95% simultaneous confidence intervals to determine which mean components differ among the populations represented by the three time periods. Are the usual MANOVA assumptions realistic for these data? Explain.

6.25. Construct a one-way MANOVA of the crude-oil data listed in Table 11.7 on page 662. Construct 95% simultaneous confidence intervals to determine which mean components differ among the populations. (You may want to consider transformations of the data to make them more closely conform to the usual MANOVA assumptions.)

Table 6.12 Oxygen-Consumption Data

Males				Females			
x_1 Resting O_2 (L/min)	x_2 Resting O_2 (mL/kg/min)	x_3 Maximum O_2 (L/min)	x_4 Maximum O_2 (mL/kg/min)	x_1 Resting O_2 (L/min)	x_2 Resting O_2 (mL/kg/min)	x_3 Maximum O_2 (L/min)	x_4 Maximum O_2 (mL/kg/min)
0.34	3.71	2.87	30.87	0.29	5.04	1.93	33.85
0.39	5.08	3.38	43.85	0.28	3.95	2.51	35.82
0.48	5.13	4.13	44.51	0.31	4.88	2.31	36.40
0.31	3.95	3.60	46.00	0.30	5.97	1.90	37.87
0.36	5.51	3.11	47.02	0.28	4.57	2.32	38.30
0.33	4.07	3.95	48.50	0.11	1.74	2.49	39.19
0.43	4.77	4.39	48.75	0.25	4.66	2.12	39.21
0.48	6.69	3.50	48.86	0.26	5.28	1.98	39.94
0.21	3.71	2.82	48.92	0.39	7.32	2.25	42.41
0.32	4.35	3.59	48.38	0.37	6.22	1.71	28.97
0.54	7.89	3.47	50.56	0.31	4.20	2.76	37.80
0.32	5.37	3.07	51.15	0.35	5.10	2.10	31.10
0.40	4.95	4.43	55.34	0.29	4.46	2.50	38.30
0.31	4.97	3.56	56.67	0.33	5.60	3.06	51.80
0.44	6.68	3.86	58.49	0.18	2.80	2.40	37.60
0.32	4.80	3.31	49.99	0.28	4.01	2.58	36.78
0.50	6.43	3.29	42.25	0.44	6.69	3.05	46.16
0.36	5.99	3.10	51.70	0.22	4.55	1.85	38.95
0.48	6.30	4.80	63.30	0.34	5.73	2.43	40.60
0.40	6.00	3.06	46.23	0.30	5.12	2.58	43.69
0.42	6.04	3.85	55.08	0.31	4.77	1.97	30.40
0.55	6.45	5.00	58.80	0.27	5.16	2.03	39.46
0.50	5.55	5.23	57.46	0.66	11.05	2.32	39.34
0.34	4.27	4.00	50.35	0.37	5.23	2.48	34.86
0.40	4.58	2.82	32.48	0.35	5.37	2.25	35.07

Source: Data courtesy of S. Rokicki.

Table 6.13 Egyptian Skull Data

MaxBreath (x_1)	BasHeight (x_2)	BasLength (x_3)	NasHeight (x_4)	Time Period
131	138	89	49	1
125	131	92	48	1
131	132	99	50	1
119	132	96	44	1
136	143	100	54	1
138	137	89	56	1
139	130	108	48	1
125	136	93	48	1
131	134	102	51	1
134	134	99	51	1
⋮	⋮	⋮	⋮	⋮
124	138	101	48	2
133	134	97	48	2
138	134	98	45	2
148	129	104	51	2
126	124	95	45	2
135	136	98	52	2
132	145	100	54	2
133	130	102	48	2
131	134	96	50	2
133	125	94	46	2
⋮	⋮	⋮	⋮	⋮
132	130	91	52	3
133	131	100	50	3
138	137	94	51	3
130	127	99	45	3
136	133	91	49	3
134	123	95	52	3
136	137	101	54	3
133	131	96	49	3
138	133	100	55	3
138	133	91	46	3

Source: Data courtesy of J. Jackson.

6.26. A project was designed to investigate how consumers in Green Bay, Wisconsin, would react to an electrical time-of-use pricing scheme. The cost of electricity during peak periods for some customers was set at eight times the cost of electricity during off-peak hours. Hourly consumption (in kilowatt-hours) was measured on a hot summer day in July and compared, for both the test group and the control group, with baseline consumption measured on a similar day before the experimental rates began. The responses,

$$\log(\text{current consumption}) - \log(\text{baseline consumption})$$

for the hours ending 9 A.M. 11 A.M. (a peak hour), 1 P.M., and 3 P.M. (a peak hour) produced the following summary statistics:

Test group:	$n_1 = 28, \bar{\mathbf{x}}_1' = [.153, -.231, -.322, -.339]$
Control group:	$n_2 = 58, \bar{\mathbf{x}}_2' = [.151, .180, .256, .257]$
and	

$$\mathbf{S}_{\text{pooled}} = \begin{bmatrix} .804 & .355 & .228 & .232 \\ .355 & .722 & .233 & .199 \\ .228 & .233 & .592 & .239 \\ .232 & .199 & .239 & .479 \end{bmatrix}$$

Source: Data courtesy of *Statistical Laboratory*, University of Wisconsin.

Perform a profile analysis. Does time-of-use pricing seem to make a difference in electrical consumption? What is the nature of this difference, if any? Comment. (Use a significance level of $\alpha = .05$ for any statistical tests.)

6.27. As part of the study of love and marriage in Example 6.14, a sample of husbands and wives were asked to respond to these questions:

1. What is the level of passionate love you feel for your partner?
2. What is the level of passionate love that your partner feels for you?
3. What is the level of companionate love that you feel for your partner?
4. What is the level of companionate love that your partner feels for you?

The responses were recorded on the following 5-point scale.

None at all	Very little	Some	A great deal	Tremendous amount
1	2	3	4	5

Thirty husbands and 30 wives gave the responses in Table 6.14, where X_1 = a 5-point-scale response to Question 1, X_2 = a 5-point-scale response to Question 2, X_3 = a 5-point-scale response to Question 3, and X_4 = a 5-point-scale response to Question 4.

(a) Plot the mean vectors for husbands and wives as sample profiles.

(b) Is the husband rating wife profile parallel to the wife rating husband profile? Test for parallel profiles with $\alpha = .05$. If the profiles appear to be parallel, test for coincident profiles at the same level of significance. Finally, if the profiles are coincident, test for level profiles with $\alpha = .05$. What conclusion(s) can be drawn from this analysis?

6.28. Two species of biting flies (genus *Leptoconops*) are so similar morphologically, that for many years they were thought to be the same. Biological differences such as sex ratios of emerging flies and biting habits were found to exist. Do the taxonomic data listed in part in Table 6.15 on page 352 and on the website www.prenhall.com/statistics indicate any difference in the two species *L. carteri* and *L. torrens*? Test for the equality of the two population mean vectors using $\alpha = .05$. If the hypotheses of equal mean vectors is rejected, determine the mean components (or linear combinations of mean components) most responsible for rejecting H_0. Justify your use of normal-theory methods for these data.

6.29. Using the data on bone mineral content in Table 1.8, investigate equality between the dominant and nondominant bones.

Table 6.14 Spouse Data

Husband rating wife				Wife rating husband			
x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
2	3	5	5	4	4	5	5
5	5	4	4	4	5	5	5
4	5	5	5	4	4	5	5
4	3	4	4	4	5	5	5
3	3	5	5	4	4	5	5
3	3	4	5	3	3	4	4
3	4	4	4	4	3	5	4
4	4	5	5	3	4	5	5
4	5	5	5	4	4	5	4
4	4	3	3	3	4	4	4
4	4	5	5	4	5	5	5
5	5	4	4	5	5	5	5
4	4	4	4	4	4	5	5
4	3	5	5	4	4	4	4
4	4	5	5	4	4	5	5
3	3	4	5	3	4	4	4
4	5	4	4	5	5	5	5
5	5	5	5	4	5	4	4
5	5	4	4	3	4	4	4
4	4	4	4	5	3	4	4
4	4	4	4	5	3	4	4
4	4	4	4	4	5	4	4
3	4	5	5	2	5	5	5
5	3	5	5	3	4	5	5
5	5	3	3	4	3	5	5
3	3	4	4	4	4	4	4
4	4	4	4	4	4	5	5
3	3	5	5	3	4	4	4
4	4	3	3	4	4	5	4
4	4	5	5	4	4	5	5

Source: Data courtesy of E. Hatfield.

(a) Test using $\alpha = .05$.

(b) Construct 95% simultaneous confidence intervals for the mean differences.

(c) Construct the Bonferroni 95% simultaneous intervals, and compare these with the intervals in Part b.

6.30. Table 6.16 on page 353 contains the bone mineral contents, for the first 24 subjects in Table 1.8, 1 year after their participation in an experimental program. Compare the data from both tables to determine whether there has been bone loss.

(a) Test using $\alpha = .05$.

(b) Construct 95% simultaneous confidence intervals for the mean differences.

(c) Construct the Bonferroni 95% simultaneous intervals, and compare these with the intervals in Part b.

Table 6.15 Biting-Fly Data

	x_1 (Wing length)	x_2 (Wing width)	x_3 (Third palp length)	x_4 (Third palp width)	x_5 (Fourth palp length)	x_6 (Length of antennal segment 12)	x_7 (Length of antennal segment 13)
	85	41	31	13	25	9	8
	87	38	32	14	22	13	13
	94	44	36	15	27	8	9
	92	43	32	17	28	9	9
	96	43	35	14	26	10	10
	91	44	36	12	24	9	9
	90	42	36	16	26	9	9
	92	43	36	17	26	9	9
	91	41	36	14	23	9	9
	87	38	35	11	24	9	10
L. torrens	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	106	47	38	15	26	10	10
	105	46	34	14	31	10	11
	103	44	34	15	23	10	10
	100	41	35	14	24	10	10
	109	44	36	13	27	11	10
	104	45	36	15	30	10	10
	95	40	35	14	23	9	10
	104	44	34	15	29	9	10
	90	40	37	12	22	9	10
	104	46	37	14	30	10	10
	86	19	37	11	25	9	9
	94	40	38	14	31	6	7
	103	48	39	14	33	10	10
	82	41	35	12	25	9	8
	103	43	42	15	32	9	9
	101	43	40	15	25	9	9
	103	45	44	14	29	11	11
	100	43	40	18	31	11	10
	99	41	42	15	31	10	10
	100	44	43	16	34	10	10
L. carteri	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	99	42	38	14	33	9	9
	110	45	41	17	36	9	10
	99	44	35	16	31	10	10
	103	43	38	14	32	10	10
	95	46	36	15	31	8	8
	101	47	38	14	37	11	11
	103	47	40	15	32	11	11
	99	43	37	14	23	11	10
	105	50	40	16	33	12	11
	99	47	39	14	34	7	7

Source: Data courtesy of William Atchley.

Table 6.16 Mineral Content in Bones (After 1 Year)

Subject number	Dominant radius	Radius	Dominant humerus	Humerus	Dominant ulna	Ulna
1	1.027	1.051	2.268	2.246	.869	.964
2	.857	.817	1.718	1.710	.602	.689
3	.875	.880	1.953	1.756	.765	.738
4	.873	.698	1.668	1.443	.761	.698
5	.811	.813	1.643	1.661	.551	.619
6	.640	.734	1.396	1.378	.753	.515
7	.947	.865	1.851	1.686	.708	.787
8	.886	.806	1.742	1.815	.687	.715
9	.991	.923	1.931	1.776	.844	.656
10	.977	.925	1.933	2.106	.869	.789
11	.825	.826	1.609	1.651	.654	.726
12	.851	.765	2.352	1.980	.692	.526
13	.770	.730	1.470	1.420	.670	.580
14	.912	.875	1.846	1.809	.823	.773
15	.905	.826	1.842	1.579	.746	.729
16	.756	.727	1.747	1.860	.656	.506
17	.765	.764	1.923	1.941	.693	.740
18	.932	.914	2.190	1.997	.883	.785
19	.843	.782	1.242	1.228	.577	.627
20	.879	.906	2.164	1.999	.802	.769
21	.673	.537	1.573	1.330	.540	.498
22	.949	.900	2.130	2.159	.804	.779
23	.463	.637	1.041	1.265	.570	.634
24	.776	.743	1.442	1.411	.585	.640

Source: Data courtesy of Everett Smith.

6.31. Peanuts are an important crop in parts of the southern United States. In an effort to develop improved plants, crop scientists routinely compare varieties with respect to several variables. The data for one two-factor experiment are given in Table 6.17 on page 354. Three varieties (5, 6, and 8) were grown at two geographical locations (1, 2) and, in this case, the three variables representing yield and the two important grade–grain characteristics were measured. The three variables are

X_1 = Yield (plot weight)

X_2 = Sound mature kernels (weight in grams—maximum of 250 grams)

X_3 = Seed size (weight, in grams, of 100 seeds)

There were two replications of the experiment.

(a) Perform a two-factor MANOVA using the data in Table 6.17. Test for a location effect, a variety effect, and a location–variety interaction. Use $\alpha = .05$.

(b) Analyze the residuals from Part a. Do the usual MANOVA assumptions appear to be satisfied? Discuss.

(c) Using the results in Part a, can we conclude that the location and/or variety effects are additive? If not, does the interaction effect show up for some variables, but not for others? Check by running three separate univariate two-factor ANOVAs.

Table 6.17 Peanut Data

Factor 1 Location	Factor 2 Variety	x_1 Yield	x_2 SdMatKer	x_3 SeedSize
1	5	195.3	153.1	51.4
1	5	194.3	167.7	53.7
2	5	189.7	139.5	55.5
2	5	180.4	121.1	44.4
1	6	203.0	156.8	49.8
1	6	195.9	166.0	45.8
2	6	202.7	166.1	60.4
2	6	197.6	161.8	54.1
1	8	193.5	164.5	57.8
1	8	187.0	165.1	58.6
2	8	201.5	166.8	65.0
2	8	200.0	173.8	67.2

Source: Data courtesy of Yolanda Lopez.

(d) Larger numbers correspond to better yield and grade-grain characteristics. Using location 2, can we conclude that one variety is better than the other two for each characteristic? Discuss your answer, using 95% Bonferroni simultaneous intervals for pairs of varieties.

6.32. In one experiment involving remote sensing, the spectral reflectance of three species of 1-year-old seedlings was measured at various wavelengths during the growing season. The seedlings were grown with two different levels of nutrient: the optimal level, coded +, and a suboptimal level, coded −. The species of seedlings used were sitka spruce (SS), Japanese larch (JL), and lodgepole pine (LP). Two of the variables measured were

$$X_1 = \text{percent spectral reflectance at wavelength 560 nm (green)}$$

$$X_2 = \text{percent spectral reflectance at wavelength 720 nm (near infrared)}$$

The cell means (CM) for Julian day 235 for each combination of species and nutrient level are as follows. These averages are based on four replications.

560CM	720CM	Species	Nutrient
10.35	25.93	SS	+
13.41	38.63	JL	+
7.78	25.15	LP	+
10.40	24.25	SS	−
17.78	41.45	JL	−
10.40	29.20	LP	−

(a) Treating the cell means as individual observations, perform a two-way MANOVA to test for a species effect and a nutrient effect. Use $\alpha = .05$.

(b) Construct a two-way ANOVA for the 560CM observations and another two-way ANOVA for the 720CM observations. Are these results consistent with the MANOVA results in Part a? If not, can you explain any differences?

6.33. Refer to Exercise 6.32. The data in Table 6.18 are measurements on the variables

X_1 = percent spectral reflectance at wavelength 560 nm (green)
X_2 = percent spectral reflectance at wavelength 720 nm (near infrared)

for three species (sitka spruce [SS], Japanese larch [JL], and lodgepole pine [LP]) of 1-year-old seedlings taken at three different times (Julian day 150 [1], Julian day 235 [2], and Julian day 320 [3]) during the growing season. The seedlings were all grown with the optimal level of nutrient.

(a) Perform a two-factor MANOVA using the data in Table 6.18. Test for a species effect, a time effect and species–time interaction. Use $\alpha = .05$.

Table 6.18 Spectral Reflectance Data

560 nm	720 nm	Species	Time	Replication
9.33	19.14	SS	1	1
8.74	19.55	SS	1	2
9.31	19.24	SS	1	3
8.27	16.37	SS	1	4
10.22	25.00	SS	2	1
10.13	25.32	SS	2	2
10.42	27.12	SS	2	3
10.62	26.28	SS	2	4
15.25	38.89	SS	3	1
16.22	36.67	SS	3	2
17.24	40.74	SS	3	3
12.77	67.50	SS	3	4
12.07	33.03	JL	1	1
11.03	32.37	JL	1	2
12.48	31.31	JL	1	3
12.12	33.33	JL	1	4
15.38	40.00	JL	2	1
14.21	40.48	JL	2	2
9.69	33.90	JL	2	3
14.35	40.15	JL	2	4
38.71	77.14	JL	3	1
44.74	78.57	JL	3	2
36.67	71.43	JL	3	3
37.21	45.00	JL	3	4
8.73	23.27	LP	1	1
7.94	20.87	LP	1	2
8.37	22.16	LP	1	3
7.86	21.78	LP	1	4
8.45	26.32	LP	2	1
6.79	22.73	LP	2	2
8.34	26.67	LP	2	3
7.54	24.87	LP	2	4
14.04	44.44	LP	3	1
13.51	37.93	LP	3	2
13.33	37.93	LP	3	3
12.77	60.87	LP	3	4

Source: Data courtesy of Mairtin Mac Siurtain.

(b) Do you think the usual MANOVA assumptions are satisfied for the these data? Discuss with reference to a residual analysis, and the possibility of correlated observations over time.

(c) Foresters are particularly interested in the interaction of species and time. Does interaction show up for one variable but not for the other? Check by running a univariate two-factor ANOVA for each of the two responses.

(d) Can you think of another method of analyzing these data (or a different experimental design) that would allow for a potential time trend in the spectral reflectance numbers?

6.34. Refer to Example 6.15.

(a) Plot the profiles, the components of $\bar{\mathbf{x}}_1$ versus time and those of $\bar{\mathbf{x}}_2$ versus time, on the same graph. Comment on the comparison.

(b) Test that linear growth is adequate. Take $\alpha = .01$.

6.35. Refer to Example 6.15 but treat all 31 subjects as a single group. The maximum likelihood estimate of the $(q + 1) \times 1$ $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{B}'\mathbf{S}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{S}^{-1}\bar{\mathbf{x}}$$

where $\mathbf{S}$ is the sample covariance matrix.

The estimated covariances of the maximum likelihood estimators are

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}) = \frac{(n-1)(n-2)}{(n-1-p+q)(n-p+q)n}(\mathbf{B}'\mathbf{S}^{-1}\mathbf{B})^{-1}$$

Fit a quadratic growth curve to this single group and comment on the fit.

6.36. Refer to Example 6.4. Given the summary information on electrical usage in this example, use Box's M-test to test the hypothesis $H_0: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. Here $\boldsymbol{\Sigma}_1$ is the covariance matrix for the two measures of usage for the population of Wisconsin homeowners *with* air conditioning, and $\boldsymbol{\Sigma}_2$ is the electrical usage covariance matrix for the population of Wisconsin homeowners *without* air conditioning. Set $\alpha = .05$.

6.37. Table 6.9 page 344 contains the carapace measurements for 24 female and 24 male turtles. Use Box's M-test to test $H_0: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}_1$ is the population covariance matrix for carapace measurements for female turtles, and $\boldsymbol{\Sigma}_2$ is the population covariance matrix for carapace measurements for male turtles. Set $\alpha = .05$.

6.38. Table 11.7 page 662 contains the values of three trace elements and two measures of hydrocarbons for crude oil samples taken from three groups (zones) of sandstone. Use Box's M-test to test equality of population covariance matrices for the three sandstone groups. Set $\alpha = .05$. Here there are $p = 5$ variables and you may wish to consider transformations of the measurements on these variables to make them more nearly normal.

6.39. Anacondas are some of the largest snakes in the world. Jesus Ravis and his fellow researchers capture a snake and measure its (i) snout vent length (cm) or the length from the snout of the snake to its vent where it evacuates waste and (ii) weight (kilograms). A sample of these measurements in shown in Table 6.19.

(a) Test for equality of means between males and females using $\alpha = .05$. Apply the large sample statistic.

(b) Is it reasonable to pool variances in this case? Explain.

(c) Find the 95% Boneferroni confidence intervals for the mean differences between males and females on both length and weight.

Table 6.19 Anaconda Data

Snout vent Length	Weight	Gender	Snout vent length	Weight	Gender
271.0	18.50	F	176.7	3.00	M
477.0	82.50	F	259.5	9.75	M
306.3	23.40	F	258.0	10.07	M
365.3	33.50	F	229.8	7.50	M
466.0	69.00	F	233.0	6.25	M
440.7	54.00	F	237.5	9.85	M
315.0	24.97	F	268.3	10.00	M
417.5	56.75	F	222.5	9.00	M
307.3	23.15	F	186.5	3.75	M
319.0	29.51	F	238.8	9.75	M
303.9	19.98	F	257.6	9.75	M
331.7	24.00	F	172.0	3.00	M
435.0	70.37	F	244.7	10.00	M
261.3	15.50	F	224.7	7.25	M
384.8	63.00	F	231.7	9.25	M
360.3	39.00	F	235.9	7.50	M
441.4	53.00	F	236.5	5.75	M
246.7	15.75	F	247.4	7.75	M
365.3	44.00	F	223.0	5.75	M
336.8	30.00	F	223.7	5.75	M
326.7	34.00	F	212.5	7.65	M
312.0	25.00	F	223.2	7.75	M
226.7	9.25	F	225.0	5.84	M
347.4	30.00	F	228.0	7.53	M
280.2	15.25	F	215.6	5.75	M
290.7	21.50	F	221.0	6.45	M
438.6	57.00	F	236.7	6.49	M
377.1	61.50	F	235.3	6.00	M

Source: Data Courtesy of Jesus Ravis.

6.40. Compare the male national track records in Table 8.6 with the female national track records in Table 1.9 using the results for the 100m, 200m, 400m, 800m and 1500m races. Treat the data as a random sample of size 64 of the twelve record values.

(a) Test for equality of means between males and females using $\alpha = .05$. Explain why it may be appropriate to analyze differences.

(b) Find the 95% Bonferroni confidence intervals for the mean differences between male and females on all of the races.

6.41. When cell phone relay towers are not working properly, wireless providers can lose great amounts of money so it is important to be able to fix problems expeditiously. A first step toward understanding the problems involved is to collect data from a designed experiment involving three factors. A problem was initially classified as low or high severity, simple or complex, and the engineer assigned was rated as relatively new (novice) or expert (guru).

Two times were observed. The time to assess the problem and plan an attack and the time to implement the solution were each measured in hours. The data are given in Table 6.20.

Perform a MANOVA including appropriate confidence intervals for important effects.

Table 6.20 Fixing Breakdowns

Problem Severity Level	Problem Complexity Level	Engineer Experience Level	Problem Assessment Time	Problem Implementation Time	Total Resolution Time
Low	Simple	Novice	3.0	6.3	9.3
Low	Simple	Novice	2.3	5.3	7.6
Low	Simple	Guru	1.7	2.1	3.8
Low	Simple	Guru	1.2	1.6	2.8
Low	Complex	Novice	6.7	12.6	19.3
Low	Complex	Novice	7.1	12.8	19.9
Low	Complex	Guru	5.6	8.8	14.4
Low	Complex	Guru	4.5	9.2	13.7
High	Simple	Novice	4.5	9.5	14.0
High	Simple	Novice	4.7	10.7	15.4
High	Simple	Guru	3.1	6.3	9.4
High	Simple	Guru	3.0	5.6	8.6
High	Complex	Novice	7.9	15.6	23.5
High	Complex	Novice	6.9	14.9	21.8
High	Complex	Guru	5.0	10.4	15.4
High	Complex	Guru	5.3	10.4	15.7

Source: Data courtesy of Dan Porter.

References

1. Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* (3rd ed.). New York: John Wiley, 2003.
2. Bacon-Shone, J., and W. K. Fung. "A New Graphical Method for Detecting Single and Multiple Outliers in Univariate and Multivariate Data." *Applied Statistics*, **36**, no. 2 (1987), 153–162.
3. Bartlett, M. S. "Properties of Sufficiency and Statistical Tests." *Proceedings of the Royal Society of London (A)*, **160** (1937), 268–282.
4. Bartlett, M. S. "Further Aspects of the Theory of Multiple Regression." *Proceedings of the Cambridge Philosophical Society*, **34** (1938), 33–40.
5. Bartlett, M. S. "Multivariate Analysis." *Journal of the Royal Statistical Society Supplement (B)*, **9** (1947), 176–197.
6. Bartlett, M. S. "A Note on the Multiplying Factors for Various χ^2 Approximations." *Journal of the Royal Statistical Society (B)*, **16** (1954), 296–298.
7. Box, G. E. P., "A General Distribution Theory for a Class of Likelihood Criteria." *Biometrika*, **36** (1949), 317–346.
8. Box, G. E. P., "Problems in the Analysis of Growth and Wear Curves." *Biometrics*, **6** (1950), 362–389.

9. Box, G. E. P., and N. R. Draper. *Evolutionary Operation: A Statistical Method for Process Improvement*. New York: John Wiley, 1969.

10. Box, G. E. P., W. G. Hunter, and J. S. Hunter. *Statistics for Experimenters* (2nd ed.). New York: John Wiley, 2005.

11. Johnson, R. A. and G. K. Bhattacharyya. *Statistics: Principles and Methods* (5th ed.). New York: John Wiley, 2005.

12. Jolicoeur, P., and J. E. Mosimann. "Size and Shape Variation in the Painted Turtle: A Principal Component Analysis." *Growth*, **24** (1960), 339–354.

13. Khattree, R. and D. N. Naik, *Applied Multivariate Statistics with SAS® Software* (2nd ed.). Cary, NC: SAS Institute Inc., 1999.

14. Kshirsagar, A. M., and W. B. Smith, *Growth Curves*. New York: Marcel Dekker, 1995.

15. Krishnamoorthy, K., and J. Yu. "Modified Nel and Van der Merwe Test for the Multivariate Behrens-Fisher Problem." *Statistics & Probability Letters*, **66** (2004), 161–169.

16. Mardia, K. V., "The Effect of Nonnormality on some Multivariate Tests and Robustnes to Nonnormality in the Linear Model." *Biometrika*, **58** (1971), 105-121.

17. Montgomery, D. C. *Design and Analysis of Experiments* (6th ed.). New York: John Wiley, 2005.

18. Morrison, D. F. *Multivariate Statistical Methods* (4th ed.). Belmont, CA: Brooks/Cole Thomson Learning, 2005.

19. Nel, D. G., and C. A. Van der Merwe. "A Solution to the Multivariate Behrens-Fisher Problem." *Communications in Statistics—Theory and Methods*, **15** (1986), 3719–3735.

20. Pearson, E. S., and H. O. Hartley, eds. *Biometrika Tables for Statisticians*. vol. II. Cambridge, England: Cambridge University Press, 1972.

21. Potthoff, R. F. and S. N. Roy. "A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems." *Biometrika*, **51** (1964), 313–326.

22. Scheffé, H. *The Analysis of Variance*. New York: John Wiley, 1959.

23. Tiku, M. L., and N. Balakrishnan. "Testing the Equality of Variance–Covariance Matrices the Robust Way." *Communications in Statistics—Theory and Methods*, **14**, no. 12 (1985), 3033–3051.

24. Tiku, M. L., and M. Singh. "Robust Statistics for Testing Mean Vectors of Multivariate Distributions." *Communications in Statistics—Theory and Methods*, **11**, no. 9 (1982), 985–1001.

25. Wilks, S. S. "Certain Generalizations in the Analysis of Variance." *Biometrika*, **24** (1932), 471–494.