References

1. Agresti, A. *Categorical Data Analysis* (2nd ed.), New York: John Wiley, 2002.
2. Bacon-Sone, 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. Bickel, P. J., and K. A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics,* Vol. I (2nd ed.), Upper Saddle River, NJ: Prentice Hall, 2000.
4. Bishop, Y. M. M., S. E. Feinberg, and P. W. Holland. *Discrete Multivariate Analysis: Theory and Practice* (Paperback). Cambridge, MA: The MIT Press, 1977.
5. Dempster, A. P., N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm (with Discussion)." *Journal of the Royal Statistical Society* (B), **39**, no. 1 (1977), 1–38.
6. Hartley, H. O. "Maximum Likelihood Estimation from Incomplete Data." *Biometrics,* **14** (1958), 174–194.
7. Hartley, H. O., and R. R. Hocking. "The Analysis of Incomplete Data." *Biometrics,* **27** (1971), 783–808.
8. Johnson, R. A. and T. Langeland "A Linear Combinations Test for Detecting Serial Correlation in Multivariate Samples." *Topics in Statistical Dependence.* (1991) Institute of Mathematical Statistics Monograph, Eds. Block, H. et al., 299–313.
9. Johnson, R. A. and R. Li "Multivariate Statistical Process Control Schemes for Controlling a Mean." *Springer Handbook of Engineering Statistics* (2006), H. Pham, Ed. Springer, Berlin.
10. Ryan, T. P. *Statistical Methods for Quality Improvement* (2nd ed.). New York: John Wiley, 2000.
11. 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.

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

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

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 *jth unit* as

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

and the p paired-difference random variables become

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

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

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

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

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

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

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

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

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

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.)

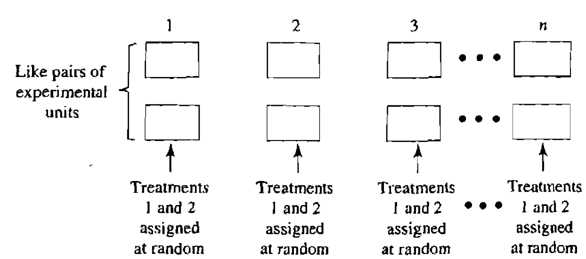
Our analysis assumed a normal distribution for the $\mathbf{D}_j$. In fact, the situation is further complicated by the presence of one or, possibly, two outliers. (See Exercise 6.3.) These data can be transformed to data more nearly normal, but with such a small sample, it is difficult to remove the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

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

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

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

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 $(\mathbf{BC}_2)'(\mathbf{BC}_2\mathbf{SC}_2'\mathbf{B}')^{-1}(\mathbf{BC}_2) = \mathbf{C}_2'\mathbf{B}'(\mathbf{B}')^{-1}(\mathbf{C}_2\mathbf{SC}_2')^{-1}\mathbf{B}^{-1}\mathbf{BC}_2 = \mathbf{C}_2'(\mathbf{C}_2\mathbf{SC}_2')^{-1}\mathbf{C}_2$, 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

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

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)

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

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:

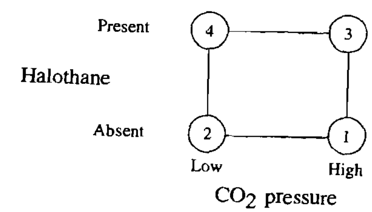


Table 6.2 contains the four measurements for each of the 19 dogs, where

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

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

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

$$(\mu_1 + \mu_4) - (\mu_2 + \mu_3) = \left(\begin{array}{c}\text{Contrast representing the influence} \\ \text{of halothane on } CO_2 \text{ pressure differences} \\ (\text{H--}CO_2 \text{ pressure “interaction”})\end{array}\right)$$

Table 6.2 Sleeping-Dog Data

Dog	Treatment 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

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

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, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ (same covariance matrix). (6-20)

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

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

We set

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

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,

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

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

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

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

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

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

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

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

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,

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

$$= \begin{pmatrix}\text{multivariate normal}\\ \text{random vector}\end{pmatrix}' \left(\frac{\text{Wishart random matrix}}{\text{d.f.}}\right)^{-1} \begin{pmatrix}\text{multivariate normal}\\ \text{random vector}\end{pmatrix}$$

$$= N_p(\mathbf{0}, \boldsymbol{\Sigma})' \left[\frac{W_{n_1+n_2-2}(\boldsymbol{\Sigma})}{n_1 + n_2 - 2}\right]^{-1} N_p(\mathbf{0}, \boldsymbol{\Sigma})$$

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

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

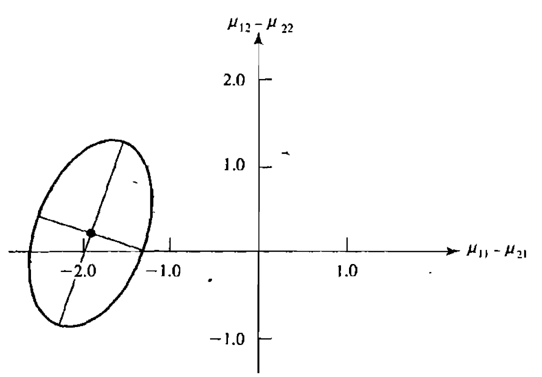


Figure 6.1 95% confidence ellipse for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.

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, $s_{1,\mathbf{a}}^2 = \mathbf{a}'\mathbf{S}_1\mathbf{a}$ and $s_{2,\mathbf{a}}^2 = \mathbf{a}'\mathbf{S}_2\mathbf{a}$

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

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

To test H_0: $\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mathbf{a}'\boldsymbol{\delta}_0$, 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

$$t_{\mathbf{a}}^2 = \frac{[\mathbf{a}'(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) - \mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s_{\mathbf{a},\text{pooled}}^2} = \frac{[\mathbf{a}'(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))]^2}{\mathbf{a}'\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\mathbf{a}} \qquad (6\text{-}26)$$

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),

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

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

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

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

Remark. For testing H_0: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$, the linear combination $\hat{\mathbf{a}}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$, with coefficient vector $\hat{\mathbf{a}} \propto \mathbf{S}_{\text{pooled}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$, 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

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

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

(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

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

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

or

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

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.

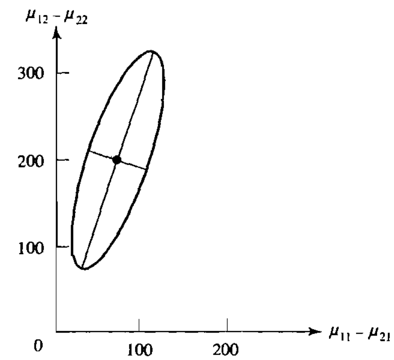


Figure 6.2 95% confidence ellipse for $\boldsymbol{\mu}_1' - \boldsymbol{\mu}_2' = (\mu_{11} - \mu_{21}, \mu_{12} - \mu_{22})$.

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

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

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),

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

and

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

By the central limit theorem, $\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2$ 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 $\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2$ to $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ would be

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

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 $H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ 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 $H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$. ■

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 $\boldsymbol{\Sigma}_1 \ne \boldsymbol{\Sigma}_2$) 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

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

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

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

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

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

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

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*

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

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

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

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,

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

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

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

Similarly,

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

and the residual sum of squares is

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

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

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

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

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

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

or

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

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

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

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

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	$SS_{tr} = \sum_{\ell=1}^{g} n_\ell(\bar{x}_\ell - \bar{x})^2$	$g - 1$
Residual (error)	$SS_{res} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)^2$	$\sum_{\ell=1}^{g} n_\ell - g$
Total (corrected for the mean)	$SS_{cor} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})^2$	$\sum_{\ell=1}^{g} n_\ell - 1$

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

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

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_{tr}/SS_{res} or for large values of $1 + SS_{tr}/SS_{res}$. The statistic appropriate for a multivariate generalization rejects H_0 for *small* values of the reciprocal

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

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

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

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,

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

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

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