whereas the univariate tests inflate α. Consequently, when the multivariate and univariate results disagree, our tendency is to trust the multivariate result. In Section 5.5, we discuss various procedures for ascertaining the contribution of the individual variables after the multivariate test has rejected the hypothesis. □

5.3 TESTS ON $\boldsymbol{\mu}$ WHEN $\boldsymbol{\Sigma}$ IS UNKNOWN

In Section 5.2, we said little about properties of the tests, because the tests discussed were of slight practical consequence due to the assumption that $\boldsymbol{\Sigma}$ is known. We will be more concerned with test properties in Sections 5.3 and 5.4, first in the one-sample case and then in the two-sample case. The reader may wonder why we include one-sample tests, since we seldom, if ever, have need of a test for $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$. However, we will cover this case for two reasons:

1. Many general principles are more easily illustrated in the one-sample framework than in the two-sample case.
2. Some very useful tests can be cast in the one-sample framework. Two examples are (1) $H_0: \boldsymbol{\mu}_d = \mathbf{0}$ used in the paired comparison test covered in Section 5.7 and (2) $H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ used in profile analysis in Section 5.9, in analysis of repeated measures in Section 6.9, and in growth curves in Section 6.10.

5.3.1 Review of Univariate *t*-Test for H_0: $\mu = \mu_0$ with σ Unknown

We first review the familiar one-sample t-test in the univariate case, with only one variable measured on each sampling unit. We assume that a random sample $y_1, y_2, \ldots, y_n$ is available from $N(\mu, \sigma^2)$. We estimate μ by $\overline{y}$ and σ^2 by s^2, where $\overline{y}$ and s^2 are given by (3.1) and (3.4). To test $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$, we use

$$t = \frac{\overline{y} - \mu_0}{s/\sqrt{n}} = \frac{\sqrt{n}(\overline{y} - \mu_0)}{s}. \tag{5.3}$$

If H_0 is true, t is distributed as t_{n-1}, where $n-1$ is the degrees of freedom. We reject H_0 if $|\sqrt{n}(\overline{y} - \mu_0)/s| \geq t_{\alpha/2,n-1}$, where $t_{\alpha/2,n-1}$ is a critical value from the t-table.

The first expression in (5.3), $t = (\overline{y} - \mu_0)/(s/\sqrt{n})$, is the *characteristic form* of the t-statistic, which represents a sample standardized distance between $\overline{y}$ and μ_0. In this form, the hypothesized mean is subtracted from $\overline{y}$ and the difference is divided by $s_{\overline{y}} = s/\sqrt{n}$. Since $y_1, y_2, \ldots, y_n$ is a random sample from $N(\mu, \sigma^2)$, the random variables $\overline{y}$ and s are independent. We will see an analogous characteristic form for the T^2-statistic in the multivariate case in Section 5.3.2.

5.3.2 Hotelling's T^2-Test for H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ with $\boldsymbol{\Sigma}$ Unknown

We now move to the multivariate case in which p variables are measured on each sampling unit. We assume that a random sample $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ is available from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\mathbf{y}_i$ contains the p measurements on the ith sampling unit (subject or object). We estimate $\boldsymbol{\mu}$ by $\overline{\mathbf{y}}$ and $\boldsymbol{\Sigma}$ by $\mathbf{S}$, where $\overline{\mathbf{y}}$ and $\mathbf{S}$ are given by (3.16), (3.19), (3.22), (3.27), and (3.29). In order to test $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1\colon \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, we use an extension of the univariate t-statistic in (5.3). In squared form, the univariate t can be rewritten as

$$t^2 = \frac{n(\overline{y} - \mu_0)^2}{s^2} = n(\overline{y} - \mu_0)(s^2)^{-1}(\overline{y} - \mu_0). \tag{5.4}$$

When $\overline{y} - \mu_0$ and s^2 are replaced by $\overline{\mathbf{y}} - \boldsymbol{\mu}_0$ and $\mathbf{S}$, we obtain the test statistic

$$T^2 = n(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{y}} - \boldsymbol{\mu}_0). \tag{5.5}$$

Alternatively, T^2 can be obtained from Z^2 in (5.2) by replacing $\boldsymbol{\Sigma}$ with $\mathbf{S}$.

The distribution of T^2 was obtained by Hotelling (1931), assuming H_0 is true and sampling is from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The distribution is indexed by two parameters, the dimension p and the degrees of freedom $\nu = n - 1$. We reject H_0 if $T^2 > T^2_{\alpha,p,n-1}$ and accept H_0 otherwise. Critical values of the T^2-distribution are found in Table A.7, taken from Kramer and Jensen (1969a).

Note that the terminology "accept H_0" is used for expositional convenience to describe our decision when we do not reject the hypothesis. Strictly speaking, we do not accept H_0 in the sense of actually believing it is true. If the sample size were extremely large and we accepted H_0, we could be reasonably certain that the true $\boldsymbol{\mu}$ is close to the hypothesized value $\boldsymbol{\mu}_0$. Otherwise, accepting H_0 means only that we have failed to reject H_0.

The T^2-statistic can be viewed as the sample standardized distance between the observed sample mean vector and the hypothetical mean vector. If the sample mean vector is notably distant from the hypothetical mean vector, we become suspicious of the hypothetical mean vector and wish to reject H_0.

The test statistic is a scalar quantity, since $T^2 = n(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)$ is a quadratic form. As with the χ^2-distribution of Z^2, the density of T^2 is skewed because the lower limit is zero and there is no upper limit.

The *characteristic form* of the T^2-statistic (5.5) is

$$T^2 = (\overline{\mathbf{y}} - \boldsymbol{\mu}_0)' \left(\frac{\mathbf{S}}{n}\right)^{-1} (\overline{\mathbf{y}} - \boldsymbol{\mu}_0). \tag{5.6}$$

The characteristic form has two features:

1. $\mathbf{S}/n$ is the sample covariance matrix of $\overline{\mathbf{y}}$ and serves as a standardizing matrix in the distance function.
2. Since $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, it follows that $\overline{\mathbf{y}}$ is $N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$, $(n-1)\mathbf{S}$ is $W(n-1, \boldsymbol{\Sigma})$, and $\overline{\mathbf{y}}$ and $\mathbf{S}$ are independent (see Section 4.3.2).

In (5.3), the univariate t-statistic represents the number of standard deviations $\overline{y}$ is separated from μ_0. In appearance, the T^2-statistic (5.6) is similar, but no such simple interpretation is possible. If we add a variable, the distance in (5.6) increases. (By analogy, the hypotenuse of a right triangle is longer than either of the legs.) Thus we need a test statistic that indicates the significance of the distance from $\overline{\mathbf{y}}$ to $\boldsymbol{\mu}_0$, while allowing for the number of dimensions (see comment 3 at the end of this section about the T^2-table). Since the resulting T^2-statistic cannot be readily interpreted in terms of the number of standard deviations $\overline{\mathbf{y}}$ is from $\boldsymbol{\mu}_0$, we do not have an intuitive feel for its significance as we do with the univariate t. We must compare the calculated value of T^2 with the table value. In addition, the T^2-table provides some insights into the behavior of the T^2-distribution. Four of these insights are noted at the end of this section.

If a test leads to rejection of $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$, the question arises as to which variable or variables contributed most to the rejection. This issue is discussed in Section 5.5 for the two-sample T^2-test of $H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, and the results there can be easily adapted to the one-sample test of $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$. For confidence intervals on the individual μ_j's in $\boldsymbol{\mu}$, see Rencher (1998, Section 3.4).

The following are some key properties of the T^2-test:

1. We must have $n - 1 > p$. Otherwise, $\mathbf{S}$ is singular and T^2 cannot be computed.
2. In both the one-sample and two-sample cases, the degrees of freedom for the T^2-statistic will be the same as for the analogous univariate t-test; that is, $\nu = n - 1$ for one sample and $\nu = n_1 + n_2 - 2$ for two samples (see Section 5.4.2).
3. The alternative hypothesis is two-sided. Because the space is multidimensional, we do not consider one-sided alternative hypotheses, such as $\boldsymbol{\mu} > \boldsymbol{\mu}_0$. However, even though the alternative hypothesis $H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ is essentially two-sided, the critical region is one-tailed (we reject H_0 for large values of T^2). This is typical of many multivariate tests.
4. In the univariate case, $t^2_{n-1} = F_{1,n-1}$. The statistic T^2 can also be converted to an F-statistic as follows:

$$\frac{\nu - p + 1}{\nu p} T^2_{p,\nu} = F_{p,\nu-p+1}. \tag{5.7}$$

   Note that the dimension p (number of variables) of the T^2-statistic becomes the first of the two degrees-of-freedom parameters of the F. The number of degrees of freedom for T^2 is denoted by ν, and the F transformation is given in terms of a general ν, since other applications of T^2 will have ν different from $n - 1$ (see, for example, Sections 5.4.2 and 6.3.2).

Equation (5.7) gives an easy way to find critical values for the T^2-test. However, we have provided critical values of T^2 in Table A.7 because of the insights they provide into the behavior of the T^2-distribution in particular and multivariate tests in general. The following are some insights that can readily be gleaned from the T^2-tables:

1. The first column of Table A.7 contains squares of t-table values; that is, $T^2_{\alpha,1,\nu} = t^2_{\alpha/2,\nu}$. (We use $t^2_{\alpha/2}$ because the univariate test of $H_0\colon \mu = \mu_0$ vs. $H_1\colon \mu \neq \mu_0$ is two-tailed.) Thus for $p = 1$, T^2 reduces to t^2. This can easily be seen by comparing (5.5) with (5.4).
2. The last row of each page of Table A.7 contains χ^2 critical values, that is, $T^2_{p,\infty} = \chi^2_p$. Thus as n increases, $\mathbf{S}$ approaches $\boldsymbol{\Sigma}$, and

$$T^2 = n(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)$$

approaches $Z^2 = n(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)$ in (5.2), which is distributed as χ^2_p.
3. The values increase along each row of Table A.7; that is, for a fixed ν, the critical value $T^2_{\alpha,p,\nu}$ increases with p. It was noted above that in any given sample, the calculated value of T^2 increases if a variable is added. However, since the critical value also increases, a variable should not be added unless it adds a significant amount to T^2.
4. As p increases, larger values of ν are required for the distribution of T^2 to approach χ^2. In the univariate case, t in (5.3) is considered a good approximation to the standard normal z in (5.1) when $\nu = n - 1$ is at least 30. In the first column ($p = 1$) of Table A.7, we see $T^2_{.05,1,30} = 4.171$ and $T^2_{.05,1,\infty} = 3.841$, with a ratio of $4.171/3.841 = 1.086$. For $p = 5$, ν must be 100 to obtain the same ratio: $T^2_{.05,5,100}/T^2_{.05,5,\infty} = 1.086$. For $p = 10$, we need $\nu = 200$ to obtain a similar value of the ratio: $T^2_{.05,10,200}/T^2_{.05,10,\infty} = 1.076$. Thus one must be very cautious in stating that T^2 has an approximate χ^2-distribution for large n. The α level (Type I error rate) could be substantially inflated. For example, suppose $p = 10$ and we assume that $n = 30$ is sufficiently large for a χ^2-approximation to hold. Then we would reject H_0 for $T^2 \geq 18.307$ with a target α-level of .05. However, the correct critical value is 34.044, and the misuse of 18.307 would yield an actual α of $P(T^2_{10,29} \geq 18.307) = .314$.

Example 5.3.2. In Table 3.3 we have $n = 10$ observations on $p = 3$ variables. Desirable levels for y_1 and y_2 are 15.0 and 6.0, respectively, and the expected level of y_3 is 2.85. We can, therefore, test the hypothesis

$$H_0\colon \boldsymbol{\mu} = \begin{pmatrix} 15.0 \\ 6.0 \\ 2.85 \end{pmatrix}.$$

In Examples 3.5 and 3.6, $\bar{\mathbf{y}}$ and $\mathbf{S}$ were obtained as

$$\bar{\mathbf{y}} = \begin{pmatrix} 28.1 \\ 7.18 \\ 3.09 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 140.54 & 49.68 & 1.94 \\ 49.68 & 72.25 & 3.68 \\ 1.94 & 3.68 & .25 \end{pmatrix}.$$

To test H_0, we use (5.5):

$$
\begin{aligned}
T^2 &= n(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{y}} - \boldsymbol{\mu}_0) \\
&= 10 \begin{pmatrix} 28.1 & - & 15.0 \\ 7.18 & - & 6.0 \\ 3.09 & - & 2.85 \end{pmatrix}' \begin{pmatrix} 140.54 & 49.68 & 1.94 \\ 49.68 & 72.25 & 3.68 \\ 1.94 & 3.68 & .25 \end{pmatrix}^{-1} \begin{pmatrix} 28.1 & - & 15.0 \\ 7.18 & - & 6.0 \\ 3.09 & - & 2.85 \end{pmatrix} \\
&= 24.559.
\end{aligned}
$$

From Table A.7, we obtain the critical value $T^2_{.05,3,9} = 16.766$. Since the observed value of T^2 exceeds the critical value, we reject the hypothesis. □

5.4 COMPARING TWO MEAN VECTORS

We first review the univariate two-sample t-test and then proceed with the analogous multivariate test.

5.4.1 Review of Univariate Two-Sample *t*-Test

In the one-variable case we obtain a random sample $y_{11}, y_{12}, \ldots, y_{1n_1}$ from $N(\mu_1, \sigma_1^2)$ and a second random sample $y_{21}, y_{22}, \ldots, y_{2n_2}$ from $N(\mu_2, \sigma_2^2)$. We assume that the two samples are independent and that $\sigma_1^2 = \sigma_2^2 = \sigma^2$, say, with σ^2 unknown. [The assumptions of independence and equal variances are necessary in order for the t-statistic in (5.8) to have a t-distribution.] From the two samples we calculate $\overline{y}_1$, $\overline{y}_2$, $\text{SS}_1 = \sum_{i=1}^{n_1}(y_{1i} - \overline{y}_1)^2 = (n_1 - 1)s_1^2$, $\text{SS}_2 = \sum_{i=1}^{n_2}(y_{2i} - \overline{y}_2)^2 = (n_2 - 1)s_2^2$, and the pooled variance

$$
s_{\text{pl}}^2 = \frac{\text{SS}_1 + \text{SS}_2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},
$$

where $n_1 + n_2 - 2$ is the sum of the weights $n_1 - 1$ and $n_2 - 1$ in the numerator. With this denominator, s_{pl}^2 is an unbiased estimator for the common variance, σ^2, that is, $E(s_{\text{pl}}^2) = \sigma^2$.

To test

$$
H_0\colon \mu_1 = \mu_2 \quad \text{vs.} \quad H_1\colon \mu_1 \neq \mu_2,
$$

we use

$$
t = \frac{\overline{y}_1 - \overline{y}_2}{s_{\text{pl}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}, \tag{5.8}
$$

which has a t-distribution with $n_1 + n_2 - 2$ degrees of freedom when H_0 is true. We therefore reject H_0 if $|t| \geq t_{\alpha/2, n_1+n_2-2}$.

Note that (5.8) exhibits the *characteristic form* of a t-statistic. In this form, the denominator is the sample standard deviation of the numerator; that is,

$$s_{\text{pl}}\sqrt{1/n_1 + 1/n_2}$$

is an estimate of

$$\sigma_{\overline{y}_1 - \overline{y}_2} = \sqrt{\text{var}(\overline{y}_1 - \overline{y}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
$$= \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

5.4.2 Multivariate Two-Sample T^2-Test

We now consider the case where p variables are measured on each sampling unit in two samples. We wish to test

$$H_0\colon \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs.} \quad H_1\colon \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

We obtain a random sample $\mathbf{y}_{11}, \mathbf{y}_{12}, \ldots, \mathbf{y}_{1n_1}$ from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and a second random sample $\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2n_2}$ from $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$. We assume that the two samples are independent and that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, say, with $\boldsymbol{\Sigma}$ unknown. These assumptions are necessary in order for the T^2-statistic in (5.9) to have a T^2-distribution. A test of $H_0\colon \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ is given in Section 7.3.2. For an approximate test of $H_0\colon \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ that can be used when $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, see Rencher (1998, Section 3.9).

The sample mean vectors are $\overline{\mathbf{y}}_1 = \sum_{i=1}^{n_1} \mathbf{y}_{1i}/n_1$ and $\overline{\mathbf{y}}_2 = \sum_{i=1}^{n_2} \mathbf{y}_{2i}/n_2$. Define $\mathbf{W}_1$ and $\mathbf{W}_2$ to be the matrices of sums of squares and cross products for the two samples:

$$\mathbf{W}_1 = \sum_{i=1}^{n_1} (\mathbf{y}_{1i} - \overline{\mathbf{y}}_1)(\mathbf{y}_{1i} - \overline{\mathbf{y}}_1)' = (n_1 - 1)\mathbf{S}_1,$$

$$\mathbf{W}_2 = \sum_{i=1}^{n_2} (\mathbf{y}_{2i} - \overline{\mathbf{y}}_2)(\mathbf{y}_{2i} - \overline{\mathbf{y}}_2)' = (n_2 - 1)\mathbf{S}_2.$$

Since $(n_1 - 1)\mathbf{S}_1$ is an unbiased estimator of $(n_1 - 1)\boldsymbol{\Sigma}$ and $(n_2 - 1)\mathbf{S}_2$ is an unbiased estimator of $(n_2 - 1)\boldsymbol{\Sigma}$, we can pool them to obtain an unbiased estimator of the common population covariance matrix, $\boldsymbol{\Sigma}$:

$$\mathbf{S}_{\text{pl}} = \frac{1}{n_1 + n_2 - 2}(\mathbf{W}_1 + \mathbf{W}_2)$$

$$= \frac{1}{n_1 + n_2 - 2}[(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2].$$

Thus $E(\mathbf{S}_{\text{pl}}) = \mathbf{\Sigma}$.

The square of the univariate t-statistic (5.8) can be expressed as

$$t^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{y}_1 - \overline{y}_2)(s_{\text{pl}}^2)^{-1}(\overline{y}_1 - \overline{y}_2).$$

This can be generalized to p variables by substituting $\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2$ for $\overline{y}_1 - \overline{y}_2$ and $\mathbf{S}_{\text{pl}}$ for s_{pl}^2 to obtain

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)'\mathbf{S}_{\text{pl}}^{-1}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2), \tag{5.9}$$

which is distributed as T^2_{p,n_1+n_2-2} when $H_0\colon \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ is true. To carry out the test, we collect the two samples, calculate T^2 by (5.9), and reject H_0 if $T^2 \geq T^2_{\alpha,p,n_1+n_2-2}$. Critical values of T^2 are found in Table A.7. For tables of the power of the T^2-test (probability of rejecting H_0 when it is false) and illustrations of their use, see Rencher (1998, Section 3.10).

The T^2-statistic (5.9) can be expressed in *characteristic form* as the standardized distance between $\overline{\mathbf{y}}_1$ and $\overline{\mathbf{y}}_2$:

$$T^2 = (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)'\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pl}}\right]^{-1}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2), \tag{5.10}$$

where $(1/n_1 + 1/n_2)\mathbf{S}_{\text{pl}}$ is the sample covariance matrix for $\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2$ and $\mathbf{S}_{\text{pl}}$ is independent of $\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2$ because of sampling from the multivariate normal. For a discussion of robustness of T^2 to departures from the assumptions of multivariate normality and homogeneity of covariance matrices ($\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$), see Rencher (1998, Section 3.7).

Some key properties of the two-sample T^2-test are given in the following list:

1. It is necessary that $n_1 + n_2 - 2 > p$ for $\mathbf{S}_{\text{pl}}$ to be nonsingular.
2. The statistic T^2 is, of course, a scalar. The $3p + p(p-1)/2$ quantities in $\overline{\mathbf{y}}_1$, $\overline{\mathbf{y}}_2$, and $\mathbf{S}_{\text{pl}}$ have been reduced to a single scale on which T^2 is large if the sample evidence favors $H_1\colon \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ and small if the evidence supports $H_0\colon \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$; we reject H_0 if the standardized distance between $\overline{\mathbf{y}}_1$ and $\overline{\mathbf{y}}_2$ is large.
3. Since the lower limit of T^2 is zero and there is no upper limit, the density is skewed. In fact, as noted in (5.11), T^2 is directly related to F, which is a well-known skewed distribution.
4. For degrees of freedom of T^2 we have $n_1 + n_2 - 2$, which is the same as for the corresponding univariate t-statistic (5.8).

5. The alternative hypothesis $H_1: \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ is two sided. The critical region $T^2 > T^2_\alpha$ is one-tailed, however, as is typical of many multivariate tests.
6. The T^2-statistic can be readily transformed to an F-statistic using (5.7):

$$\frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2 = F_{p, n_1+n_2-p-1}, \tag{5.11}$$

where again the dimension p of the T^2-statistic becomes the first degree-of-freedom parameter for the F-statistic.

Example 5.4.2. Four psychological tests were given to 32 men and 32 women. The data are recorded in Table 5.1 (Beall 1945). The variables are

y_1 = pictorial inconsistencies $\quad$ y_3 = tool recognition

y_2 = paper form board $\quad$ y_4 = vocabulary

The mean vectors and covariance matrices of the two samples are

$$\bar{\mathbf{y}}_1 = \begin{pmatrix} 15.97 \\ 15.91 \\ 27.19 \\ 22.75 \end{pmatrix}, \qquad \bar{\mathbf{y}}_2 = \begin{pmatrix} 12.34 \\ 13.91 \\ 16.66 \\ 21.94 \end{pmatrix},$$

$$\mathbf{S}_1 = \begin{pmatrix} 5.192 & 4.545 & 6.522 & 5.250 \\ 4.545 & 13.18 & 6.760 & 6.266 \\ 6.522 & 6.760 & 28.67 & 14.47 \\ 5.250 & 6.266 & 14.47 & 16.65 \end{pmatrix},$$

$$\mathbf{S}_2 = \begin{pmatrix} 9.136 & 7.549 & 4.864 & 4.151 \\ 7.549 & 18.60 & 10.22 & 5.446 \\ 4.864 & 10.22 & 30.04 & 13.49 \\ 4.151 & 5.446 & 13.49 & 28.00 \end{pmatrix}.$$

The sample covariance matrices do not appear to indicate a disparity in the population covariance matrices. (A significance test to check this assumption is carried out in Example 7.3.2, and the hypothesis $H_0: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ is not rejected.) The pooled covariance matrix is

$$\mathbf{S}_{\text{pl}} = \frac{1}{32 + 32 - 2}[(32 - 1)\mathbf{S}_1 + (32 - 1)\mathbf{S}_2]$$
$$= \begin{pmatrix} 7.164 & 6.047 & 5.693 & 4.701 \\ 6.047 & 15.89 & 8.492 & 5.856 \\ 5.693 & 8.492 & 29.36 & 13.98 \\ 4.701 & 5.856 & 13.98 & 22.32 \end{pmatrix}.$$

Table 5.1. Four Psychological Test Scores on 32 Males and 32 Females

Males				Females			
y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4
15	17	24	14	13	14	12	21
17	15	32	26	14	12	14	26
15	14	29	23	12	19	21	21
13	12	10	16	12	13	10	16
20	17	26	28	11	20	16	16
15	21	26	21	12	9	14	18
15	13	26	22	10	13	18	24
13	5	22	22	10	8	13	23
14	7	30	17	12	20	19	23
17	15	30	27	11	10	11	27
17	17	26	20	12	18	25	25
17	20	28	24	14	18	13	26
15	15	29	24	14	10	25	28
18	19	32	28	13	16	8	14
18	18	31	27	14	8	13	25
15	14	26	21	13	16	23	28
18	17	33	26	16	21	26	26
10	14	19	17	14	17	14	14
18	21	30	29	16	16	15	23
18	21	34	26	13	16	23	24
13	17	30	24	2	6	16	21
16	16	16	16	14	16	22	26
11	15	25	23	17	17	22	28
16	13	26	16	16	13	16	14
16	13	23	21	15	14	20	26
18	18	34	24	12	10	12	9
16	15	28	27	14	17	24	23
15	16	29	24	13	15	18	20
18	19	32	23	11	16	18	28
18	16	33	23	7	7	19	18
17	20	21	21	12	15	7	28
19	19	30	28	6	5	6	13

By (5.9), we obtain

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)'\mathbf{S}_{\text{pl}}^{-1}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2) = 97.6015.$$

From interpolation in Table A.7, we obtain $T^2_{.01,4,62} = 15.373$, and we therefore reject $H_0\colon \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. See Example 5.5 for a discussion of which variables contribute most to separation of the two groups. □

Thus LR is small when t^2 is large, and rejection of H_0 for LR $\leq c$ is equivalent to rejection of H_0 for $t^2 \geq t^2_{\alpha/2}$. The t-test in (3.2) is therefore a likelihood ratio test statistic. □

The likelihood ratio method of test construction usually leads to tests that are relatively powerful and sometimes produces tests with optimum power over a wide class of alternatives. Many multivariate tests are derived in this fashion.

3.3.3 One-Sample T^2-Test

We now consider the multivariate case in which the hypothesis $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is p-dimensional. In order to test $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, we assume that a random sample $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ is available from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ unknown. By analogy with the univariate t-statistic in (3.2) and the Z^2-statistic in (3.1), we use the sample mean vector $\overline{\mathbf{y}}$ and the sample covariance matrix $\mathbf{S}$ to construct the test statistic,

$$T^2 = n(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{y}} - \boldsymbol{\mu}_0), \tag{3.14}$$

which is the standardized distance from $\overline{\mathbf{y}}$ to $\boldsymbol{\mu}_0$. We show in Section 3.3.7 that this is the likelihood ratio test statistic.

If H_0 is true and if sampling is from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then T^2 in (3.14) has Hotelling's (1931) T^2-distribution with dimension p and degrees of freedom $n - 1$. We reject H_0 if $T^2 \geq T^2_{\alpha,p,n-1}$ and accept H_0 otherwise. We use the terminology "accept H_0" for expositional convenience; it means only that we have failed to reject H_0. Critical values $T^2_{\alpha,p,n-1}$ of the T^2-distribution are given in Table B.1 in Appendix B (Kramer and Jensen 1969a). For an analysis of patterns in the T^2-table and the insights that these patterns provide about the T^2-test and multivariate testing in general, see Rencher (1995, Section 5.3.2).

For $p = 1$, the T^2-statistic in (3.14) reduces to the square of the univariate t in (3.2):

$$T^2 = n(\overline{y} - \mu_0)(s^2)^{-1}(\overline{y} - \mu_0) = \frac{n(\overline{y} - \mu_0)^2}{s^2} = t^2.$$

Another link between T^2 and the univariate t is that in cases where there is an analogous t-test, the degrees of freedom for the T^2-test will be the same as for the univariate t-test. Thus the one-sample T^2-test has $n - 1$ degrees of freedom, and the two-sample T^2-test (to be defined in Section 3.5.2) has $n_1 + n_2 - 2$ degrees of freedom.

To avoid inverting $\mathbf{S}$, an alternative formula for computing T^2 in (3.14) can be obtained using (A.7.10):

$$T^2 = \frac{|\mathbf{S} + n(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)'|}{|\mathbf{S}|} - 1. \tag{3.15}$$

A key assumption in the T^2-distribution is the independence of $\bar{\mathbf{y}}$ and $\mathbf{S}$; this assumption holds when sampling from a multivariate normal population (Theorem 2.3D). Usually $\bar{\mathbf{y}}$ and $\mathbf{S}$ are obtained from the same sample, although this is not necessary. As long as $\mathbf{S}$ is an unbiased estimator of $\boldsymbol{\Sigma}$ and is independent of $\bar{\mathbf{y}}$, the estimator $\mathbf{S}$ could come wholly or partly from another sample. For example, suppose $\bar{\mathbf{y}}_1$ and $\mathbf{S}_1$ arise from a sample of size n_1 from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and that a supplementary estimate $\mathbf{S}_2$ is available from a sample of size n_2 from $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, where the two distributions have a common covariance matrix $\boldsymbol{\Sigma}$ and the two samples are independent. We estimate $\boldsymbol{\Sigma}$ by the pooled estimator $\mathbf{S}_{\text{pl}} = [(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2]/(n_1 + n_2 - 2)$, which has $n_1 + n_2 - 2$ degrees of freedom. Then a T^2-statistic could be based on $\mathbf{S}_1, \mathbf{S}_2$, or $\mathbf{S}_{\text{pl}}$:

$$
\begin{aligned}
n_1(\bar{\mathbf{y}}_1 - \boldsymbol{\mu}_1)'\mathbf{S}_1^{-1}(\bar{\mathbf{y}}_1 - \boldsymbol{\mu}_1) &\text{ is } T^2_{p,n_1-1}, \\
n_1(\bar{\mathbf{y}}_1 - \boldsymbol{\mu}_1)'\mathbf{S}_2^{-1}(\bar{\mathbf{y}}_1 - \boldsymbol{\mu}_1) &\text{ is } T^2_{p,n_2-1}, \\
n_1(\bar{\mathbf{y}}_1 - \boldsymbol{\mu}_1)'\mathbf{S}_{\text{pl}}^{-1}(\bar{\mathbf{y}}_1 - \boldsymbol{\mu}_1) &\text{ is } T^2_{p,n_1+n_2-2}.
\end{aligned}
$$

The coefficient of the quadratic form in all three cases is n_1 because each of $\mathbf{S}_1/n_1, \mathbf{S}_2/n_1$, and $\mathbf{S}_{\text{pl}}/n_1$ estimates $\text{cov}(\bar{\mathbf{y}}_1) = \boldsymbol{\Sigma}/n_1$. Thus the leading coefficient in these T^2-statistics is the sample size for $\bar{\mathbf{y}}$, and the degrees of freedom is the denominator of the unbiased estimator of $\boldsymbol{\Sigma}$.

3.3.4 Formal Definition of T^2 and Relationship to F

The *formal definition* of a T^2 random variable is similar to the formal definition of the t random variable given in (3.3). Let $\mathbf{z}$ be distributed as the multivariate normal $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ and $\mathbf{W}$ be distributed as the Wishart $W_p(\nu, \boldsymbol{\Sigma})$, with $\mathbf{z}$ and $\mathbf{W}$ independent. Then the T^2 random variable with dimension p and degrees of freedom ν is defined as

$$
T^2 = \mathbf{z}'\left(\frac{\mathbf{W}}{\nu}\right)^{-1}\mathbf{z}. \tag{3.16}
$$

The distribution of Hotelling's T^2 can be derived from this definition.

It is easy to show that the T^2-statistic (3.14) satisfies the formal definition (3.16). Define $\bar{\mathbf{v}} = \sqrt{n}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)$ and $\mathbf{W} = (n-1)\mathbf{S}$. Then $\bar{\mathbf{v}}$ is $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ if $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, $\mathbf{W}$ is $W_p(n-1, \boldsymbol{\Sigma})$, and $\bar{\mathbf{v}}$ and $\mathbf{W}$ are independent. Hence $T^2 = \bar{\mathbf{v}}'[\mathbf{W}/(n-1)]^{-1}\bar{\mathbf{v}}$ satisfies (3.16) and can be expressed as

$$
\begin{aligned}
T^2 &= \bar{\mathbf{v}}'\left(\frac{\mathbf{W}}{n-1}\right)^{-1}\bar{\mathbf{v}} \\
&= [\sqrt{n}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)]'\left[\frac{(n-1)\mathbf{S}}{n-1}\right]^{-1}[\sqrt{n}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)] \\
&= n(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0),
\end{aligned}
$$

which is (3.14).

Table 3.1. Calcium in Soil and Turnip Greens

Observation Number	y_1	y_2	y_3
1	35	3.5	2.80
2	35	4.9	2.70
3	40	30.0	4.38
4	10	2.8	3.21
5	6	2.7	2.73
6	20	2.8	2.81
7	35	4.6	2.88
8	35	10.9	2.90
9	35	8.0	3.28
10	30	1.6	3.20

Example 3.3.5. Table 3.1 lists observations of three types of calcium measurements in soil and turnip greens (Kramer and Jensen 1969a).

Target values of these three variables are 15.0, 6.0, and 2.85. Using $\boldsymbol{\mu}_0 = (15.0, 6.0, 2.85)'$ in (3.14) gives $T_{y,x}^2 = 24.559$. We now examine the effect of each variable on T^2 by using (3.19) and (3.20). We first consider the effect of y_3 as it is added to T_y^2 based on y_1 and y_2. With $x = y_3$ we have

$$\begin{pmatrix} \bar{\mathbf{y}} \\ \bar{x} \end{pmatrix} = \left(\begin{array}{c} 28.1 \\ 7.18 \\ \hline 3.089 \end{array} \right), \qquad \begin{pmatrix} \mathbf{S}_{yy} & \mathbf{s}_{xy} \\ \mathbf{s}_{xy}' & s_x^2 \end{pmatrix} = \left(\begin{array}{cc|c} 140.54 & 49.68 & 1.94 \\ 49.68 & 72.25 & 3.68 \\ \hline 1.94 & 3.68 & .25 \end{array} \right).$$

The value of T^2 based on $\mathbf{y} = (y_1, y_2)'$ is

$$\begin{aligned} T_y^2 &= n(\bar{\mathbf{y}} - \boldsymbol{\mu}_{0y})'\mathbf{S}_{yy}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}_{0y}) \\ &= 10\begin{pmatrix} 28.1 - 15.0 \\ 7.18 - 6.0 \end{pmatrix}' \begin{pmatrix} 140.54 & 49.68 \\ 49.68 & 72.25 \end{pmatrix}^{-1} \begin{pmatrix} 28.1 - 15.0 \\ 7.18 - 6.0 \end{pmatrix} \\ &= 14.388. \end{aligned}$$

Thus, without y_3, T^2 falls from $T_{y,x}^2 = 24.559$ to $T_y^2 = 14.388$, a reduction of 10.171. To see what factors contribute to this difference, we examine the elements of the second term on the right side of both (3.19) and (3.20):

$$\hat{\boldsymbol{\beta}} = \mathbf{S}_{yy}^{-1}\mathbf{s}_{xy} = \begin{pmatrix} 140.54 & 49.68 \\ 49.68 & 72.25 \end{pmatrix}^{-1} \begin{pmatrix} 1.94 \\ 3.68 \end{pmatrix} = \begin{pmatrix} -.00551 \\ .05467 \end{pmatrix},$$

$$\hat{t}_x = \frac{\hat{\boldsymbol{\beta}}'(\bar{\mathbf{y}} - \boldsymbol{\mu}_{0y})}{s_x/\sqrt{n}} = \frac{-.007720}{\sqrt{.2501/10}} = -.0488,$$

covariance matrix of $\mathbf{y}$ by $\mathbf{S}_{yy}$ and the sample mean and hypothesized mean of $\mathbf{y}$ by $\bar{\mathbf{y}}$ and $\boldsymbol{\mu}_{0y}$, respectively. Then the effect of x on T^2 is given in the following theorem.

Theorem 3.3C. For the $p + 1$ variables $(y_1, y_2, \ldots, y_p, x) = (\mathbf{y}', x)$, T^2 can be expressed as

$$T^2_{y,x} = T^2_y + \frac{n[\hat{\boldsymbol{\beta}}'(\bar{\mathbf{y}} - \boldsymbol{\mu}_{0y}) - (\bar{x} - \mu_{0x})]^2}{s_x^2(1 - R^2)} \tag{3.19}$$

$$= T^2_y + \frac{(\hat{t}_x - t_x)^2}{1 - R^2}, \tag{3.20}$$

where $T^2_{y,x}$ is the value of T^2 based on the y's and x, T^2_y is the value of T^2 based on the y's alone, $\hat{\boldsymbol{\beta}} = \mathbf{S}_{yy}^{-1}\mathbf{s}_{xy}$ is the vector of regression coefficients of x with the y's [corrected for their means, see (7.26)], $R^2 = \mathbf{s}'_{xy}\mathbf{S}_{yy}^{-1}\mathbf{s}_{xy}/s_x^2$ is the squared multiple correlation of x regressed on the y's [see (7.66)], $t_x = \sqrt{n}(\bar{x} - \mu_{0x})/s_x$, and

$$\hat{t}_x = \frac{\hat{\boldsymbol{\beta}}'(\bar{\mathbf{y}} - \boldsymbol{\mu}_{0y})}{s_x/\sqrt{n}}. \qquad \square$$

Thus t_x is the ordinary t-statistic for x by itself, and $\hat{t}_x$ can be interpreted as a "predicted" value of t_x based on the information about $\bar{x} - \mu_{0x}$ already available in the y's. If t_x and $\hat{t}_x$ are of the same sign, there are three ways in which the contribution of x can be important: (a) t_x substantially larger in absolute value than $\hat{t}_x$, (b) $\hat{t}_x$ substantially larger in absolute value than t_x, and (c) R^2 large. Otherwise, if t_x is close to $\hat{t}_x$, so that most of the evidence $\bar{x}$ provides against the hypothesis is predictable from $\bar{\mathbf{y}}$, there is little reason to include x. If t_x and $\hat{t}_x$ are of opposite signs, their effect combines to increase T^2. Theorem 3.3C also demonstrates that if x were orthogonal to the y's ($\hat{\boldsymbol{\beta}} = \mathbf{0}$), the addition of x to T^2_y would reduce to t_x^2.

Note that (3.19) proves that the addition of a variable can only increase T^2. It may seem surprising that this increase in T^2 is inversely related to $1 - R^2$ rather than to R^2; that is, the larger the value of R^2, the larger the increase in T^2. Perhaps we can draw an analogy to simple linear regression, in which a given difference between y and $\hat{y}$ is more important if the squared correlation r^2 is larger.

The net effect of a variable on T^2 is given by the second term on the right side of (3.19) or (3.20). This effect can be either greater or less than what would be expected from its univariate contribution. It is intuitively obvious that overlap with other variables can render a variable partially redundant so that its multivariate contribution is less than its univariate effect, but heretofore it has not been easy to grasp how the contribution of a variable can be enhanced in the presence of the others. [For illustrations of such situations, see Flury (1989) and Hamilton (1987).] In Theorem 3.3C, the breakdown of the effect of each variable makes clear how this can happen. Note the linearity inherent in the effect of each variable, as manifested by the presence of $\hat{\boldsymbol{\beta}}$ and R^2.

The square of a univariate t has an F-distribution. In the multivariate case, a simple function of T^2 also has an F-distribution, as shown in the following theorem.

Theorem 3.3B. The T^2-statistic with ν degrees of freedom can be transformed to an F-statistic with p and $\nu - p + 1$ degrees of freedom:

$$\frac{\nu - p + 1}{\nu p} T^2_{p,\nu} = F_{p,\nu-p+1}. \tag{3.17}$$

Proof. An F random variable is defined as the ratio of two independent χ^2 random variables, each divided by its degrees of freedom. To express (3.16) in this form, multiply and divide by $\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}$ to obtain

$$T^2 = \nu\mathbf{z}'\mathbf{W}^{-1}\mathbf{z} = \frac{\nu\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}}{\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}/\mathbf{z}'\mathbf{W}^{-1}\mathbf{z}}, \tag{3.18}$$

where $\mathbf{z}$ is $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, $\mathbf{W}$ is $W_p(\nu, \boldsymbol{\Sigma})$, and $\mathbf{z}$ and $\mathbf{W}$ are independent. By Theorem 2.2F, the quadratic form $\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}$ in the numerator is distributed as χ^2_p. It can be shown (Seber 1984, pp. 30–31; Styan 1989) that the denominator $\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}/\mathbf{z}'\mathbf{W}^{-1}\mathbf{z}$ is distributed as $\chi^2_{\nu-p+1}$ and is independent of $\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}$. If each of these two independent χ^2 random variables is divided by its degrees of freedom, the resulting ratio will have an F-distribution. Multiplying both sides of (3.18) by the ratio $(\nu - p + 1)/\nu p$, we obtain

$$\frac{\nu - p + 1}{\nu p} T^2 = \frac{\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}/p}{(\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}/\mathbf{z}'\mathbf{W}^{-1}\mathbf{z})/(\nu - p + 1)},$$

which, by definition, has an F-distribution with p and $\nu - p + 1$ degrees of freedom. □

3.3.5 Effect on T^2 of Adding a Variable

The addition of a variable to T^2 may either strengthen the evidence against the hypothesis or weaken it. For example, from Table B.1, we obtain, for 20 degrees of freedom,

$$T^2_{.05,6,20} - T^2_{.05,5,20} = 22.324 - 17.828 = 4.496.$$

Thus, if a sixth variable is added to the five already present, the critical value is increased by 4.496. If the new variable does not potentially increase the calculated T^2 by that amount, then T^2 is less likely to reject H_0. (A test of the significance of the increase in T^2 is given in Section 3.11.4.)

Rencher (1993) has given a breakdown of the factors that influence the increase in T^2 caused by an additional variable. The term "additional variable" is for convenience. In some cases, new variables may be available, but typically we are interested in the effect on T^2 of each of the present variables. Let x designate the variable of interest or additional variable to be added to $\mathbf{y} = (y_1, y_2, \ldots, y_p)'$. We denote the sample mean and variance of x by $\bar{x}$ and s_x^2, the vector of sample covariances of x with the y's by $\mathbf{s}_{xy}$, and the hypothesized mean of x by μ_{0x}. For consistency, we denote the sample

$$t_x = \frac{\bar{x} - \mu_{0x}}{s_x/\sqrt{n}} = \frac{3.089 - 2.85}{\sqrt{.2501/10}} = 1.511,$$

$$R^2 = \frac{\mathbf{s}'_{xy}\mathbf{S}^{-1}_{yy}\mathbf{s}_{xy}}{s_x^2} = .7607,$$

$$T^2_{y,x} - T^2_y = \frac{(\hat{t}_x - t_x)^2}{1 - R^2} = \frac{(-.0488 - 1.511)^2}{1 - .7607} = 10.171.$$

Thus the increase in T^2 due to $x = y_3$ is largely induced by $t_x = 1.511$ and the fairly high squared multiple correlation of y_3 with y_1 and y_2, $R^2 = .7607$. For $x = y_1$ and $x = y_2$, we have

x	$\hat{t}_x$	t_x	R^2	$T^2_{y,x} - T^2_y$
y_2	2.009	.438	.797	12.198
y_1	−.227	3.494	.282	19.289

For y_2, the increase in T^2 is due to $\hat{t}_x$ and R^2. For y_1, the increase is due almost entirely to t_x.

3.3.6 Properties of the T^2-Test

Some important properties of the T^2-test of $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1\colon \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ are as follows:

1. The T^2-statistic is invariant to transformations of the form $\mathbf{z}_i = \mathbf{A}\mathbf{y}_i + \mathbf{b}$, where $\mathbf{A}$ is a nonsingular matrix, that is, $T_z^2 = T_y^2$ (see problems 3.9 and 3.37). Invariance of this type, often referred to as *affine* invariance (or equivariance), includes changes in scale, as, for example, from inches to centimeters.
2. The T^2-test is the uniformly most powerful invariant test (Anderson 1984, Section 5.6.1).
3. The T^2-test is sensitive to certain departures from normality (see Section 3.7.2).
4. The T^2-test is the likelihood ratio test (see Section 3.3.7).
5. The T^2-test is the union–intersection test (see Section 3.3.8).

In the univariate case ($p = 1$), the t-statistic can be adapted to serve for a one-sided alternative hypothesis such as $H_1\colon \mu > \mu_0$. The resulting one-tailed t-test has some optimal properties. In the multivariate case, however, the T^2-test cannot be similarly adapted to have optimal properties for a one-sided alternative. For $p > 1$, the one-sided alternative $H_1\colon \boldsymbol{\mu} > \boldsymbol{\mu}_0$ can be defined to mean that $\mu_j > \mu_{0j}$ for all $j = 1, 2, \ldots, p$. Kariya and Cohen (1992) showed that for this case there is no scale invariant test statistic with satisfactory properties. The T^2-test is invariant but cannot be recommended for obvious reasons—it would reject H_0 when we want to accept it, namely, when some or all $\bar{y}_j$ are considerably less than the corresponding μ_{0j}. For additional discussion of the one-sided multivariate problem, see Perlman (1969), Marden (1982), and Troendle (1996).

Hence, in the case of H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, both the likelihood ratio and union–intersection approaches lead to the same test.

From (3.32), it is clear that any multiple of $\mathbf{a} = \mathbf{S}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)$ maximizes $t^2(\mathbf{a})$, that is, maximally separates $\mathbf{a}'\bar{\mathbf{y}}$ from $\mathbf{a}'\boldsymbol{\mu}_0$. The linear function $z = \mathbf{a}'\mathbf{y}$, with coefficient vector $\mathbf{a} = \mathbf{S}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)$, is called the *discriminant function*. We will discuss the discriminant function $z = \mathbf{a}'\mathbf{y}$ further in Sections 3.4.5, 3.5.5, 5.2, 5.4, 5.5.1, 5.7.2a, and 5.11.1.

3.4 CONFIDENCE INTERVALS AND TESTS FOR LINEAR FUNCTIONS OF $\boldsymbol{\mu}$

We now consider confidence intervals and tests for various linear combinations of $\boldsymbol{\mu}$, including the individual elements μ_j. We begin with a confidence region for the entire mean vector $\boldsymbol{\mu}$.

3.4.1 Confidence Region for $\boldsymbol{\mu}$

Since $n(\bar{\mathbf{y}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu})$ is distributed as T^2, we can make the probability statement $P[n(\bar{\mathbf{y}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}) \le T^2_{\alpha,p,n-1}] = 1 - \alpha$, from which a $100(1 - \alpha)\%$ *confidence region* for $\boldsymbol{\mu}$ is given by all vectors $\boldsymbol{\mu}$ that satisfy

$$n(\bar{\mathbf{y}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}) \le T^2_{\alpha,p,n-1}. \tag{3.34}$$

This hyperellipsoidal region for $\boldsymbol{\mu}$ is centered at $\boldsymbol{\mu} = \bar{\mathbf{y}}$. However, the values of $\boldsymbol{\mu}$ that satisfy (3.34) are not easy to visualize except in the case $p = 2$, where we can draw an ellipse. For $p > 2$, we can substitute various values of $\boldsymbol{\mu}$ into (3.34) to determine if they are inside the region. But this is equivalent to finding those values of $\boldsymbol{\mu}_0$ that would not be rejected by the T^2-test of H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ in (3.14). Thus we are back to the hypothesis test, and (3.34) provides little additional insight into the possible position of $\boldsymbol{\mu}$.

3.4.2 Confidence Interval for a Single Linear Combination $\mathbf{a}'\boldsymbol{\mu}$

By (3.28), a $100(1 - \alpha)\%$ confidence interval for $\mathbf{a}'\boldsymbol{\mu}$ is given by

$$\mathbf{a}'\bar{\mathbf{y}} - t_{\alpha/2,n-1}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}} \le \mathbf{a}'\boldsymbol{\mu} \le \mathbf{a}'\bar{\mathbf{y}} + t_{\alpha/2,n-1}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}. \tag{3.35}$$

If confidence intervals are desired for several linear combinations, see Sections 3.4.3 and 3.4.4.

3.4.3 Simultaneous Confidence Intervals for μ_j and $\mathbf{a}'\boldsymbol{\mu}$

Because the confidence region for $\boldsymbol{\mu}$ in (3.34) is unwieldy, we look for confidence intervals for μ_j or for $\mathbf{a}'\boldsymbol{\mu}$ for arbitrary $\mathbf{a}$. The linear combination $\mathbf{a}'\boldsymbol{\mu}$ allows for contrasts of the form $\mu_1 - \mu_2$ or $\mu_1 - 2\mu_2 + \mu_3$ and also yields each μ_j by choosing

$$\begin{aligned}
\overline{y}_1 - T_{\alpha,p,n-1}\sqrt{\frac{s_{11}}{n}} &< \mu_1 < \overline{y}_1 + T_{\alpha,p,n-1}\sqrt{\frac{s_{11}}{n}}, \\
\overline{y}_2 - T_{\alpha,p,n-1}\sqrt{\frac{s_{22}}{n}} &< \mu_2 < \overline{y}_2 + T_{\alpha,p,n-1}\sqrt{\frac{s_{22}}{n}}, \\
&\vdots \\
\overline{y}_p - T_{\alpha,p,n-1}\sqrt{\frac{s_{pp}}{n}} &< \mu_p < \overline{y}_p + T_{\alpha,p,n-1}\sqrt{\frac{s_{pp}}{n}}.
\end{aligned} \tag{3.37}$$

However, because (3.36) allows for all possible $\mathbf{a}'\boldsymbol{\mu}$, the intervals in (3.37) are ordinarily much wider than necessary to provide an overall confidence level of $100(1-\alpha)\%$ for these p intervals. The coefficient $T_{\alpha,p,n-1}$ is considerably greater than $t_{\alpha/2,n-1}$ used in (3.35) (except for $p = 1$). For example, with $n = 25$ and $p = 10$, we have $T_{.05,10,24} = 6.380$ and $t_{.025,24} = 2.064$. But clearly the use of p intervals of the form $\overline{y}_j \pm t_{\alpha/2,n-1}\sqrt{s_{jj}/n}$, $j = 1, 2, \ldots, p$, would be inappropriate because the overall confidence level for all p intervals would be less than $100(1-\alpha)\%$. Some coefficient between $t_{\alpha/2,n-1}$ and $T_{\alpha,p,n-1}$ is needed in order for the p intervals to achieve an overall confidence level closer to the nominal $100(1-\alpha)\%$. We will discuss such a coefficient in Section 3.4.4.

Note that when $p = 1$, the vector $\overline{\mathbf{y}}$ has only the single element $\overline{y}$, and the intervals in (3.37) reduce to

$$\overline{y} - T_{\alpha,1,n-1}\sqrt{\frac{s^2}{n}} \leq \mu \leq \overline{y} + T_{\alpha,1,n-1}\sqrt{\frac{s^2}{n}}.$$

This is the usual univariate confidence interval, since $T_{\alpha,1,n-1} = t_{\alpha/2,n-1}$.

Example 3.4.3. We illustrate the computation of simultaneous confidence intervals for the calcium data in Table 3.1. The mean vector and covariance matrix are

$$\overline{\mathbf{y}} = \begin{pmatrix} 28.1 \\ 7.18 \\ 3.09 \end{pmatrix}, \qquad \mathbf{S} = \begin{pmatrix} 140.54 & 49.68 & 1.94 \\ 49.68 & 72.25 & 3.68 \\ 1.94 & 3.68 & .25 \end{pmatrix}.$$

With $p = 3$ and $n = 10$, we obtain $T_{.05,3,9} = \sqrt{16.776} = 4.0946$ from Table B.1 in Appendix B. From (3.37), the intervals are as follows:

$$\overline{y}_1 \pm T_{.05,3,9}\sqrt{\frac{s_{11}}{n}}, \quad 28.1 \pm 4.0946\sqrt{\frac{140.54}{10}}, \quad 28.1 \pm 15.35, \quad (12.75, 43.45);$$

$$\overline{y}_2 \pm T_{.05,3,9}\sqrt{\frac{s_{22}}{n}}, \quad 7.18 \pm 4.0946\sqrt{\frac{72.25}{10}}, \quad 7.18 \pm 11.009, \quad (-3.829, 18.189);$$

$$\overline{y}_3 \pm T_{.05,3,9}\sqrt{\frac{s_{33}}{n}}, \quad 3.09 \pm 4.0946\sqrt{\frac{.25}{10}}, \quad 3.09 \pm .648, \quad (2.442, 3.738).$$

… to the large coefficient 4.0946