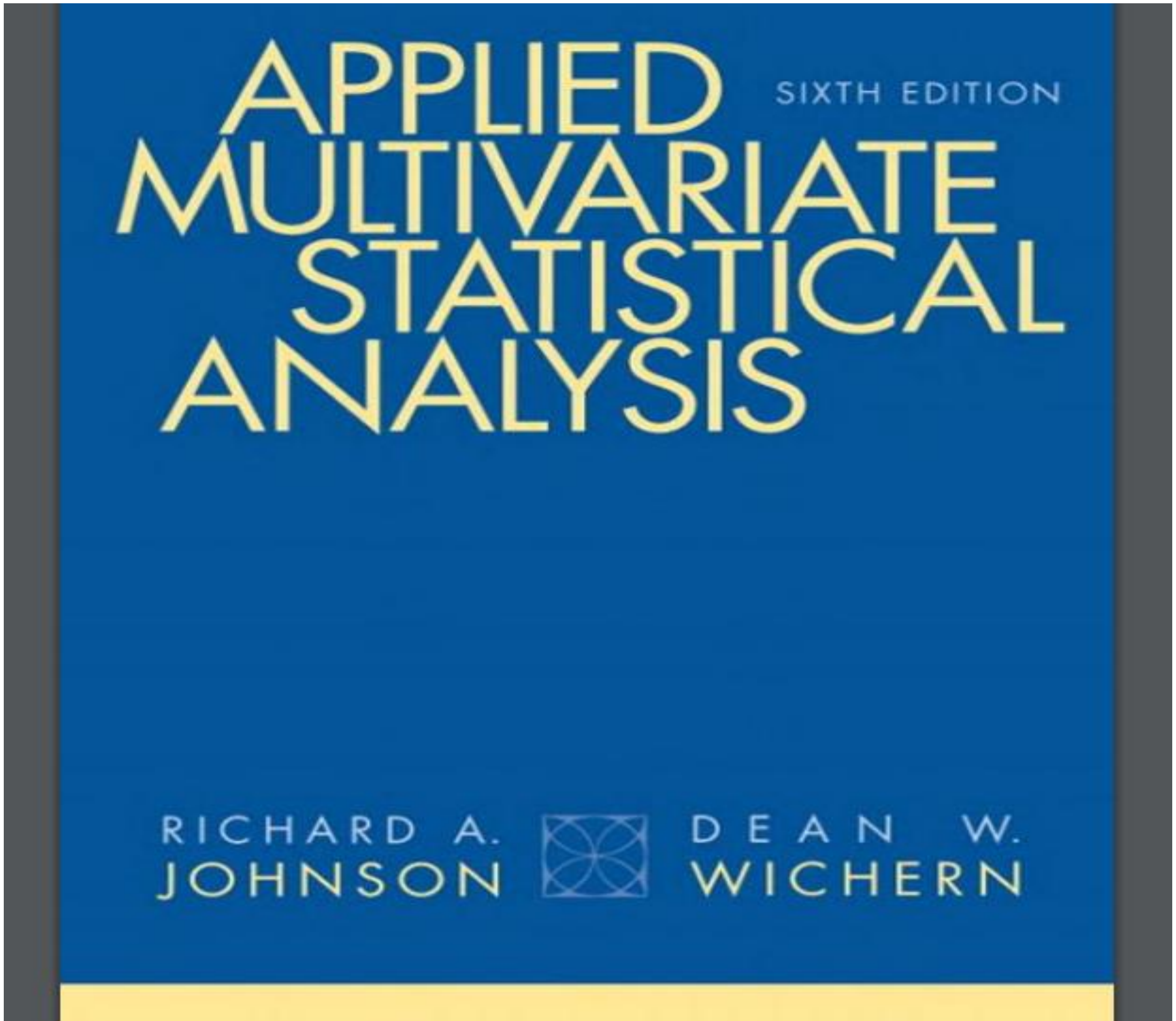
Advanced Multivariate Methods



Inferences About a Mean Vector

INFERENCES ABOUT A MEAN VECTOR

5.1 Introduction

This chapter is the first of the methodological sections of the book. We shall now use the concepts and results set forth in Chapters 1 through 4 to develop techniques for analyzing data. A large part of any analysis is concerned with *inference*—that is, reaching valid conclusions concerning a population on the basis of information from a sample.

At this point, we shall concentrate on inferences about a population mean vector and its component parts. Although we introduce statistical inference through initial discussions of tests of hypotheses, our ultimate aim is to present a full statistical analysis of the component means based on simultaneous confidence statements.

One of the central messages of multivariate analysis is that p correlated variables must be analyzed jointly. This principle is exemplified by the methods presented in this chapter.

Using μ_o As Value for Normality

5.2 The Plausibility of μ_0 as a Value for a Normal Population Mean

Let us start by recalling the univariate theory for determining whether a specific value μ_0 is a plausible value for the population mean μ. From the point of view of hypothesis testing, this problem can be formulated as a *test* of the competing *hypotheses*

$$H_0: \mu = \mu_0 \quad \text{and} \quad H_1: \mu \neq \mu_0$$

Here H_0 is the null hypothesis and H_1 is the (two-sided) alternative hypothesis. If $X_1, X_2, \ldots, X_n$ denote a random sample from a normal population, the appropriate test statistic is

$$t = \frac{(\overline{X} - \mu_0)}{s/\sqrt{n}}, \quad \text{where} \quad \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \quad \text{and} \quad s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2$$

Review of t-Test Statistics and Units of Distance as Estimated Standard Deviation

This test statistic has a student's t-distribution with $n - 1$ degrees of freedom (d.f.). We reject H_0, that μ_0 is a plausible value of μ, if the observed $|t|$ exceeds a specified percentage point of a t-distribution with $n - 1$ d.f.

Rejecting H_0 when $|t|$ is large is equivalent to rejecting H_0 if its square,

$$t^2 = \frac{(\overline{X} - \mu_0)^2}{s^2/n} = n(\overline{X} - \mu_0)(s^2)^{-1}(\overline{X} - \mu_0) \qquad (5\text{-}1)$$

is large. The variable t^2 in (5-1) is the square of the distance from the sample mean $\overline{X}$ to the test value μ_0. The units of distance are expressed in terms of $s/\sqrt{n}$, or estimated standard deviations of $\overline{X}$. Once $\overline{X}$ and s^2 are observed, the test becomes: Reject H_0 in favor of H_1, at significance level α, if

$$n(\bar{x} - \mu_0)(s^2)^{-1}(\bar{x} - \mu_0) > t^2_{n-1}(\alpha/2) \qquad (5\text{-}2)$$

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

Hypothesis Testing and Confidence Interval Using μ_o

If H_0 is not rejected, we conclude that μ_0 is a plausible value for the normal population mean. Are there other values of μ which are also consistent with the data? The answer is yes! In fact, there is always a *set* of plausible values for a normal population mean. From the well-known correspondence between acceptance regions for tests of $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ and confidence intervals for μ, we have

$$\{\text{Do not reject } H_0: \mu = \mu_0 \text{ at level } \alpha\} \quad \text{or} \quad \left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right| \leq t_{n-1}(\alpha/2)$$

is equivalent to

$$\left\{\mu_0 \text{ lies in the } 100(1 - \alpha)\% \text{ confidence interval } \bar{x} \pm t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}\right\}$$

or

$$\bar{x} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}} \qquad \text{(5-3)}$$

The confidence interval consists of all those values μ_0 that would not be rejected by the level α test of $H_0: \mu = \mu_0$.

Confidence Interval Univariate Level vs. Multivariate Analog, i.e. Squared Distance

Before the sample is selected, the $100(1-\alpha)\%$ confidence interval in (5-3) is a *random interval* because the endpoints depend upon the random variables $\overline{X}$ and s. The probability that the interval contains μ is $1-\alpha$; among large numbers of such independent intervals, approximately $100(1-\alpha)\%$ of them will contain μ.

Consider now the problem of determining whether a given $p \times 1$ vector $\boldsymbol{\mu}_0$ is a plausible value for the mean of a multivariate normal distribution. We shall proceed by analogy to the univariate development just presented.

A natural generalization of the squared distance in (5-1) is its multivariate analog

$$T^2 = (\overline{\mathbf{X}} - \boldsymbol{\mu}_0)'\left(\frac{1}{n}\mathbf{S}\right)^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}_0) = n(\overline{\mathbf{X}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}_0) \qquad (5\text{-}4)$$

Multivariate Analog Distribution: Hotelling's T^2

where

$$\underset{(p\times 1)}{\overline{\mathbf{X}}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j, \qquad \underset{(p\times p)}{\mathbf{S}} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}_j-\overline{\mathbf{X}})(\mathbf{X}_j-\overline{\mathbf{X}})', \text{ and } \underset{(p\times 1)}{\boldsymbol{\mu}_0} = \begin{bmatrix}\mu_{10}\\ \mu_{20}\\ \vdots \\ \mu_{p0}\end{bmatrix}$$

The statistic T^2 is called *Hotelling's* T^2 in honor of Harold Hotelling, a pioneer in multivariate analysis, who first obtained its sampling distribution. Here $(1/n)\mathbf{S}$ is the estimated covariance matrix of $\overline{\mathbf{X}}$. (See Result 3.1.)

If the observed statistical distance T^2 is too large—that is, if $\bar{\mathbf{x}}$ is "too far" from $\boldsymbol{\mu}_0$—the hypothesis H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ is rejected. It turns out that special tables of T^2 percentage points are not required for formal tests of hypotheses. This is true because

$$T^2 \text{ is distributed as } \frac{(n-1)p}{(n-p)}F_{p,n-p} \tag{5-5}$$

where $F_{p,n-p}$ denotes a random variable with an F-distribution with p and $n-p$ d.f.

To summarize, we have the following:

F Distribution and Hotelling's T^2

To summarize, we have the following:

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Then with $\overline{\mathbf{X}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{X}_j$ and $\mathbf{S} = \frac{1}{(n-1)} \sum_{j=1}^{n} (\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$,

$$\alpha = P\left[T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right]$$
$$= P\left[n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right] \tag{5-6}$$

whatever the true $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Here $F_{p,n-p}(\alpha)$ is the upper (100α)th percentile of the $F_{p,n-p}$ distribution.

Statement (5-6) leads immediately to a test of the hypothesis $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1\colon \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. At the α level of significance, we reject H_0 in favor of H_1 if the observed

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha) \tag{5-7}$$

F Distribution and Hotelling's T^2 Based on Wishart Assumption of the Generalized Chi-Square Distribution

It is informative to discuss the nature of the T^2-distribution briefly and its correspondence with the univariate test statistic. In Section 4.4, we described the manner in which the Wishart distribution generalizes the chi-square distribution. We can write

$$T^2 = \sqrt{n}\,(\overline{\mathbf{X}} - \boldsymbol{\mu}_0)' \left(\frac{\sum_{j=1}^{n} (\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'}{n-1} \right)^{-1} \sqrt{n}\,(\overline{\mathbf{X}} - \boldsymbol{\mu}_0)$$

Area Under Curve (Volume) and Hotelling's T^2

which combines a normal, $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, random vector and a Wishart, $\mathbf{W}_{p,n-1}(\boldsymbol{\Sigma})$, random matrix in the form

$$T^2_{p,n-1} = \begin{pmatrix} \text{multivariate normal} \\ \text{random vector} \end{pmatrix}' \left(\frac{\begin{matrix}\text{Wishart random} \\ \text{matrix}\end{matrix}}{\text{d.f.}} \right)^{-1} \begin{pmatrix} \text{multivariate normal} \\ \text{random vector} \end{pmatrix}$$

$$= N_p(\mathbf{0}, \boldsymbol{\Sigma})' \left[\frac{1}{n-1} \mathbf{W}_{p,n-1}(\boldsymbol{\Sigma}) \right]^{-1} N_p(\mathbf{0}, \boldsymbol{\Sigma}) \qquad (5\text{-}8)$$

This is analogous to

$$t^2 = \sqrt{n}\,(\overline{X} - \mu_0)(s^2)^{-1}\sqrt{n}\,(\overline{X} - \mu_0)$$

or

$$t^2_{n-1} = \begin{pmatrix} \text{normal} \\ \text{random variable} \end{pmatrix} \left(\frac{\begin{matrix}\text{(scaled) chi-square} \\ \text{random variable}\end{matrix}}{\text{d.f.}} \right)^{-1} \begin{pmatrix} \text{normal} \\ \text{random variable} \end{pmatrix}$$

for the univariate case. Since the multivariate normal and Wishart random variables are independently distributed [see (4-23)], their joint density function is the product of the marginal normal and Wishart distributions. Using calculus, the distribution (5-5) of T^2 as given previously can be derived from this joint distribution and the representation (5-8).

Evaluating Hotelling's T^2

Example 5.1 (Evaluating T^2) Let the data matrix for a random sample of size $n = 3$ from a bivariate normal population be

$$\mathbf{X} = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}$$

Evaluate the observed T^2 for $\boldsymbol{\mu}_0' = [9, 5]$. What is the sampling distribution of T^2 in this case? We find

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} \dfrac{6 + 10 + 8}{3} \\ \dfrac{9 + 6 + 3}{3} \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$$

and

$$s_{11} = \frac{(6-8)^2 + (10-8)^2 + (8-8)^2}{2} = 4$$

$$s_{12} = \frac{(6-8)(9-6) + (10-8)(6-6) + (8-8)(3-6)}{2} = -3$$

$$s_{22} = \frac{(9-6)^2 + (6-6)^2 + (3-6)^2}{2} = 9$$

Finding Hotelling's T^2

so

$$\mathbf{S} = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}$$

Thus,

$$\mathbf{S}^{-1} = \frac{1}{(4)(9) - (-3)(-3)} \begin{bmatrix} 9 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{27} \end{bmatrix}$$

and, from (5-4),

$$T^2 = 3[8 - 9, \quad 6 - 5] \begin{bmatrix} \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{27} \end{bmatrix} \begin{bmatrix} 8 - 9 \\ 6 - 5 \end{bmatrix} = 3[-1, \quad 1] \begin{bmatrix} -\frac{2}{9} \\ \frac{1}{27} \end{bmatrix} = \frac{7}{9}$$

Before the sample is selected, T^2 has the distribution of a

$$\frac{(3 - 1)2}{(3 - 2)} F_{2,3-2} = 4F_{2,1}$$

random variable.

Testing Multivariate Mean Vector with Hotelling's T^2

The next example illustrates a test of the hypothesis H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ using data collected as part of a search for new diagnostic techniques at the University of Wisconsin Medical School.

Example 5.2 (Testing a multivariate mean vector with T^2) Perspiration from 20 healthy females was analyzed. Three components, X_1 = sweat rate, X_2 = sodium content, and X_3 = potassium content, were measured, and the results, which we call the *sweat data*, are presented in Table 5.1.

Test the hypothesis H_0: $\boldsymbol{\mu}' = [4, 50, 10]$ against H_1: $\boldsymbol{\mu}' \neq [4, 50, 10]$ at level of significance $\alpha = .10$.

Computer calculations provide

$$\bar{\mathbf{x}} = \begin{bmatrix} 4.640 \\ 45.400 \\ 9.965 \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} 2.879 & 10.010 & -1.810 \\ 10.010 & 199.788 & -5.640 \\ -1.810 & -5.640 & 3.628 \end{bmatrix}$$

and

$$\mathbf{S}^{-1} = \begin{bmatrix} .586 & -.022 & .258 \\ -.022 & .006 & -.002 \\ .258 & -.002 & .402 \end{bmatrix}$$

We evaluate

$$T^2 =$$

$$20[4.640 - 4, \;\; 45.400 - 50, \;\; 9.965 - 10] \begin{bmatrix} .586 & -.022 & .258 \\ -.022 & .006 & -.002 \\ .258 & -.002 & .402 \end{bmatrix} \begin{bmatrix} 4.640 - 4 \\ 45.400 - 50 \\ 9.965 - 10 \end{bmatrix}$$

$$= 20[.640, \;\; -4.600, \;\; -.035] \begin{bmatrix} .467 \\ -.042 \\ .160 \end{bmatrix} = 9.74$$

Example F Distribution and Hotelling's T^2

Table 5.1 Sweat Data

Individual	X_1 (Sweat rate)	X_2 (Sodium)	X_3 (Potassium)
1	3.7	48.5	9.3
2	5.7	65.1	8.0
3	3.8	47.2	10.9
4	3.2	53.2	12.0
5	3.1	55.5	9.7
6	4.6	36.1	7.9
7	2.4	24.8	14.0
8	7.2	33.1	7.6
9	6.7	47.4	8.5
10	5.4	54.1	11.3
11	3.9	36.9	12.7
12	4.5	58.8	12.3
13	3.5	27.8	9.8
14	4.5	40.2	8.4
15	1.5	13.5	10.1
16	8.5	56.4	7.1
17	4.5	71.6	8.2
18	6.5	52.8	10.9
19	4.1	44.1	11.2
20	5.5	40.9	9.4

Source: Courtesy of Dr. Gerald Bargman.

Comparing the observed $T^2 = 9.74$ with the critical value

$$\frac{(n-1)p}{(n-p)} F_{p,n-p}(.10) = \frac{19(3)}{17} F_{3,17}(.10) = 3.353(2.44) = 8.18$$

we see that $T^2 = 9.74 > 8.18$, and consequently, we reject H_0 at the 10% level of significance.

F Distribution Still Assumes Normality Based on Q-Q Plots and Hotelling's T^2

We note that H_0 will be rejected if one or more of the component means, or some combination of means, differs too much from the hypothesized values [4, 50, 10]. At this point, we have no idea which of these hypothesized values may not be supported by the data.

We have assumed that the sweat data are multivariate normal. The Q–Q plots constructed from the marginal distributions of X_1, X_2, and X_3 all approximate straight lines. Moreover, scatter plots for pairs of observations have approximate elliptical shapes, and we conclude that the normality assumption was reasonable in this case. (See Exercise 5.4.) ■

One feature of the T^2-statistic is that it is invariant (unchanged) under changes in the units of measurements for $\mathbf{X}$ of the form

$$\underset{(p\times 1)}{\mathbf{Y}} = \underset{(p\times p)}{\mathbf{C}}\underset{(p\times 1)}{\mathbf{X}} + \underset{(p\times 1)}{\mathbf{d}}, \qquad \mathbf{C} \text{ nonsingular} \tag{5-9}$$

Pre-multiplication of Center and Scaled Quantities Yields Equation of Line Conversion of Fahrenheit to Celsius

A transformation of the observations of this kind arises when a constant b_i is subtracted from the ith variable to form $X_i - b_i$ and the result is multiplied by a constant $a_i > 0$ to get $a_i(X_i - b_i)$. Premultiplication of the *centered* and *scaled* quantities $a_i(X_i - b_i)$ by any nonsingular matrix will yield Equation (5-9). As an example, the operations involved in changing X_i to $a_i(X_i - b_i)$ correspond exactly to the process of converting temperature from a Fahrenheit to a Celsius reading.

Given observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ and the transformation in (5-9), it immediately follows from Result 3.6 that

$$\bar{\mathbf{y}} = \mathbf{C}\bar{\mathbf{x}} + \mathbf{d} \quad \text{and} \quad \mathbf{S}_{\mathbf{y}} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{y}_j - \bar{\mathbf{y}})(\mathbf{y}_j - \bar{\mathbf{y}})' = \mathbf{CSC}'$$

F Distribution, Volume and Hotelling's T^2

Moreover, by (2-24) and (2-45),

$$\boldsymbol{\mu}_{\mathbf{Y}} = E(\mathbf{Y}) = E(\mathbf{CX} + \mathbf{d}) = E(\mathbf{CX}) + E(\mathbf{d}) = \mathbf{C}\boldsymbol{\mu} + \mathbf{d}$$

Therefore, T^2 computed with the $\mathbf{y}$'s and a hypothesized value $\boldsymbol{\mu}_{\mathbf{Y},0} = \mathbf{C}\boldsymbol{\mu}_0 + \mathbf{d}$ is

$$\begin{aligned}
T^2 &= n(\bar{\mathbf{y}} - \boldsymbol{\mu}_{\mathbf{Y},0})'\mathbf{S}_{\mathbf{y}}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}_{\mathbf{Y},0}) \\
&= n(\mathbf{C}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0))'(\mathbf{CSC}')^{-1}(\mathbf{C}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)) \\
&= n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{C}'(\mathbf{CSC}')^{-1}\mathbf{C}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \\
&= n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{C}'(\mathbf{C}')^{-1}\mathbf{S}^{-1}\mathbf{C}^{-1}\mathbf{C}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)
\end{aligned}$$

The last expression is recognized as the value of T^2 computed with the $\mathbf{x}$'s.

Hotelling's T^2 and Likelihood Ratio Tests

5.3 Hotelling's T^2 and Likelihood Ratio Tests

We introduced the T^2-statistic by analogy with the univariate squared distance t^2. There is a general principle for constructing test procedures called the *likelihood ratio method*, and the T^2-statistic can be derived as the likelihood ratio test of H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$. The general theory of likelihood ratio tests is beyond the scope of this book. (See [3] for a treatment of the topic.) Likelihood ratio tests have several optimal properties for reasonably large samples, and they are particularly convenient for hypotheses formulated in terms of multivariate normal parameters.

We know from (4-18) that the maximum of the multivariate normal likelihood as $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are varied over their possible values is given by

$$\max_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2} |\hat{\boldsymbol{\Sigma}}|^{n/2}} e^{-np/2} \tag{5-10}$$

where

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

are the maximum likelihood estimates. Recall that $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are those choices for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that best explain the observed values of the random sample.

Hotelling's T^2 and Likelihood Ratio Tests

Under the hypothesis H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, the normal likelihood specializes to

$$L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left(-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_0)\right)$$

The mean $\boldsymbol{\mu}_0$ is now fixed, but $\boldsymbol{\Sigma}$ can be varied to find the value that is "most likely" to have led, with $\boldsymbol{\mu}_0$ fixed, to the observed sample. This value is obtained by maximizing $L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ with respect to $\boldsymbol{\Sigma}$.

Following the steps in (4-13), the exponent in $L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ may be written as

$$\begin{aligned} -\frac{1}{2}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_0) &= -\frac{1}{2}\sum_{j=1}^{n}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'\right] \\ &= -\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'\right)\right] \end{aligned}$$

Applying Result 4.10 with $\mathbf{B} = \sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'$ and $b = n/2$, we have

$$\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\hat{\boldsymbol{\Sigma}}_0|^{n/2}} e^{-np/2} \qquad (5\text{-}11)$$

Likelihood Ratio Formula Equivalent to Wilks Lambda

with

$$\hat{\boldsymbol{\Sigma}}_0 = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'$$

To determine whether $\boldsymbol{\mu}_0$ is a plausible value of $\boldsymbol{\mu}$, the maximum of $L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ is compared with the unrestricted maximum of $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The resulting ratio is called the *likelihood ratio statistic*.

Using Equations (5-10) and (5-11), we get

$$\text{Likelihood ratio} = \Lambda = \frac{\max\limits_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})}{\max\limits_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{n/2} \tag{5-12}$$

The equivalent statistic $\Lambda^{2/n} = |\hat{\boldsymbol{\Sigma}}|/|\hat{\boldsymbol{\Sigma}}_0|$ is called *Wilks' lambda*. If the observed value of this likelihood ratio is too small, the hypothesis H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ is unlikely to be true and is, therefore, rejected. Specifically, the likelihood ratio test of H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ against H_1: $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ rejects H_0 if

Likelihood Ratio Test Power of the Ratio of Generalized Variances

The equivalent statistic $\Lambda^{2/n} = |\hat{\boldsymbol{\Sigma}}|/|\hat{\boldsymbol{\Sigma}}_0|$ is called *Wilks' lambda*. If the observed value of this likelihood ratio is too small, the hypothesis $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is unlikely to be true and is, therefore, rejected. Specifically, the likelihood ratio test of $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ rejects H_0 if

$$\Lambda = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{n/2} = \left(\frac{\left|\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right|}{\left|\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'\right|}\right)^{n/2} < c_\alpha \tag{5-13}$$

where c_α is the lower (100α)th percentile of the distribution of Λ. (Note that the likelihood ratio test statistic is a power of the ratio of generalized variances.) Fortunately, because of the following relation between T^2 and Λ, we do not need the distribution of the latter to carry out the test.

Relationship Hotelling's T^2 and Likelihood Ratio Test or Wilks Lambda

Result 5.1. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Then the test in (5-7) based on T^2 is equivalent to the *likelihood ratio test* of $H_0\colon \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1\colon \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ because

$$\Lambda^{2/n} = \left(1 + \frac{T^2}{(n-1)}\right)^{-1}$$

Proof: Hotelling's T^2 and Likelihood Ratio Tests or Wilks Lambda

Proof. Let the $(p+1) \times (p+1)$ matrix

$$\mathbf{A} = \left[\begin{array}{c|c} \sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' & \sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \\ \hline \sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' & -1 \end{array}\right] = \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right]$$

By Exercise 4.11, $|\mathbf{A}| = |\mathbf{A}_{22}||\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}| = |\mathbf{A}_{11}||\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}|$, from which we obtain

$$(-1)\left|\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\right|$$

$$= \left|\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right| \left|-1 - n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\left(\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right)^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)\right|$$

Proof: Hotelling's T^2 and Likelihood Ratio Tests or Wilks Lambda

Since, by (4-14),

$$\sum_{j=1}^{n}(\mathbf{x}_j-\boldsymbol{\mu}_0)(\mathbf{x}_j-\boldsymbol{\mu}_0)' = \sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\boldsymbol{\mu}_0)(\mathbf{x}_j-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\boldsymbol{\mu}_0)'$$

$$= \sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})' + n(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)'$$

the foregoing equality involving determinants can be written

$$(-1)\left|\sum_{j=1}^{n}(\mathbf{x}_j-\boldsymbol{\mu}_0)(\mathbf{x}_j-\boldsymbol{\mu}_0)'\right| = \left|\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right|(-1)\left(1+\frac{T^2}{(n-1)}\right)$$

or

$$|n\hat{\boldsymbol{\Sigma}}_0| = |n\hat{\boldsymbol{\Sigma}}|\left(1+\frac{T^2}{(n-1)}\right)$$

Thus,

$$\Lambda^{2/n} = \frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|} = \left(1+\frac{T^2}{(n-1)}\right)^{-1} \qquad (5\text{-}14)$$

Here H_0 is rejected for small values of $\Lambda^{2/n}$ or, equivalently, large values of T^2. The critical values of T^2 are determined by (5-6). ■

Hotelling's T^2 and Relationship to t-test

Hotelling's T^2 distribution

A multivariate method that is the multivariate counterpart of Student's t and which also forms the basis for certain multivariate control charts is based on Hotelling's T^2 distribution, which was introduced by Hotelling (1947).

Univariate t-test for mean

Recall, from Section 1.3.5.2,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

has a t distribution provided that X is normally distributed, and can be used as long as X doesn't differ greatly from a normal distribution. If we wanted to test the hypothesis that $\mu = \mu_0$, we would then have

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

so that

$$t^2 = \frac{(\bar{x} - \mu_0)^2}{s^2/n}$$

$$= n(\bar{x} - \mu_0)(s^2)^{-1}(\bar{x} - \mu_0) \, .$$

Hotelling's T^2 and Generalization to p Variables

Generalize to p variables

When T^2 is generalized to p variables it becomes

$$T^2 = n(\bar{\mathbf{x}} - \mu_0)\mathbf{S}^{-1}(\bar{\mathbf{x}} - \mu_0),$$

with

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} \qquad \mu_0 = \begin{bmatrix} \mu_1^0 \\ \mu_2^0 \\ \vdots \\ \mu_p^0 \end{bmatrix}.$$

$\mathbf{S}^{-1}$ is the inverse of the sample variance-covariance matrix, $\mathbf{S}$, and n is the sample size upon which each $\bar{x}_i$, $i = 1, 2, \ldots, p$, is based. (The diagonal elements of $\mathbf{S}$ are the variances and the off-diagonal elements are the covariances for the p variables. This is discussed further in Section 6.5.4.3.1.)

Distribution of the Hotelling's T^2

Distribution of T^2

It is well known that when $\mu = \mu_0$

$$T^2 \sim \frac{p(n-1)}{n-p} F_{(p,\, n-p)}\,,$$

with $F_{(p,\, n-p)}$ representing the F distribution with p degrees of freedom for the numerator and $n-p$ for the denominator. Thus, if μ were specified to be μ_0, this could be tested by taking a single p-variate sample of size n, then computing T^2 and comparing it with

$$\frac{p(n-1)}{n-p} F_{\alpha\,(p,\, n-p)}$$

for a suitably chosen α.

Probability Density Function Hotelling's T^2 of two Chi-Square Distributions

Probability Density Function

The F distribution is the ratio of two chi-square distributions with degrees of freedom ν_1 and ν_2, respectively, where each chi-square has first been divided by its degrees of freedom. The formula for the probability density function of the F distribution is

$$f(x) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})(\frac{\nu_1}{\nu_2})^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1}}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})(1+\frac{\nu_1 x}{\nu_2})^{\frac{\nu_1+\nu_2}{2}}}$$

where ν_1 and ν_2 are the shape parameters and Γ is the gamma function. The formula for the gamma function is

$$\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt$$

The F-Distribution

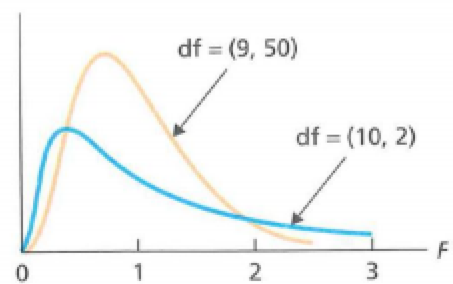
A variable is said to have an ***F*-distribution** if its distribution has the shape of a special type of right-skewed curve, called an **F-curve**. There are infinitely many F-distributions, and we identify an F-distribution (and F-curve) by its number of degrees of freedom, just as we did for t-distributions and chi-square distributions.

An F-distribution, however, has two numbers of degrees of freedom instead of one. Figure 16.1 depicts two different F-curves; one has df = (10,2), and the other has df = (9,50).

The first number of degrees of freedom for an F-curve is called the **degrees of freedom for the numerator**, and the second is called the **degrees of freedom for the denominator**. (The reason for this terminology will become clear in Section 16.3) Thus, for the F-curve in Fig. 16.1 with df = (10,2), we have

FIGURE 16.1
Two different *F*-curves

Df = (10, 2



10 = Degrees of freedom for the numerator
2 = degrees of freedom for the denominato

Key Fact 16.1 Basic Properties of F-Curves

Property 1: The total area under an F-curve equals 1
Property 2: An F-curve starts at 0 on the horizontal axis and extends indefinitely to the right, approaching, but never touching, the horizontal axis as it does so.
Property 3: An F-curve is right skewed.

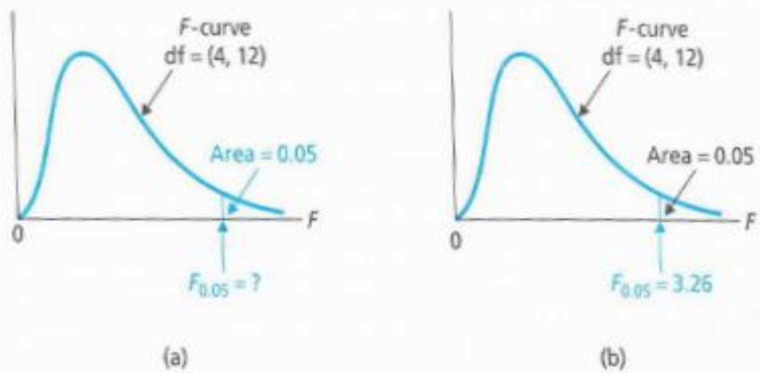
Finding F-Value Using Degrees of Freedom Numerator and Denominator

Example 16.1 Finding the F-Value Having a Specified Area to Its Right

EXAMPLE 16.1 Finding the F-Value Having a Specified Area to Its Right

For an F-curve with df = (4, 12), find $F_{0.05}$; that is, find the F-value having area 0.05 to its right, as shown in Fig. 16.2(a).

FIGURE 16.2
Finding the F-value having area 0.05 to its right



Solution We use Table VIII to find the F-value. In this case, $\alpha = 0.05$, the degrees of freedom for the numerator is 4, and the degrees of freedom for the denominator is 12.

We first go down the dfd column to "12." Next, we concentrate on the row for α labeled 0.05. Then, going across that row to the column labeled "4," we reach 3.26. This number is the F-value having area 0.05 to its right, as shown in Fig. 16.2(b). In other words, for an F-curve with df = (4, 12), $F_{0.05} = 3.26$.

You try it!

Exercise 16.7 on page 717

Hotelling's T^2 May be Calculated from Two Determinants

Incidentally, relation (5-14) shows that T^2 may be calculated from two determinants, thus avoiding the computation of $\mathbf{S}^{-1}$. Solving (5-14) for T^2, we have

$$
\begin{aligned}
T^2 &= \frac{(n-1)\,|\hat{\boldsymbol{\Sigma}}_0|}{|\hat{\boldsymbol{\Sigma}}|} - (n-1) \\
&= \frac{(n-1)\left|\sum_{j=1}^{n} (\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'\right|}{\left|\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right|} - (n-1) \qquad (5\text{-}15)
\end{aligned}
$$

Likelihood ratio tests are common in multivariate analysis. Their optimal large sample properties hold in very general contexts, as we shall indicate shortly. They are well suited for the testing situations considered in this book. Likelihood ratio methods yield test statistics that reduce to the familiar F- and t-statistics in univariate situations.

General Likelihood Ratio Method

General Likelihood Ratio Method

We shall now consider the general likelihood ratio method. Let $\boldsymbol{\theta}$ be a vector consisting of all the *unknown* population parameters, and let $L(\boldsymbol{\theta})$ be the likelihood function obtained by evaluating the joint density of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ at their observed values $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. The parameter vector $\boldsymbol{\theta}$ takes its value in the parameter set $\boldsymbol{\Theta}$. For example, in the p-dimensional multivariate normal case, $\boldsymbol{\theta}' = [\mu_1, \ldots, \mu_p, \sigma_{11}, \ldots, \sigma_{1p}, \sigma_{22}, \ldots, \sigma_{2p}, \ldots, \sigma_{p-1,p}, \sigma_{pp}]$ and $\boldsymbol{\Theta}$ consists of the p-dimensional space, where $-\infty < \mu_1 < \infty, \ldots, -\infty < \mu_p < \infty$ combined with the $[p(p+1)/2]$-dimensional space of variances and covariances such that $\boldsymbol{\Sigma}$ is positive definite. Therefore, $\boldsymbol{\Theta}$ has dimension $\nu = p + p(p+1)/2$. Under the null hypothesis H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\boldsymbol{\theta}$ is restricted to lie in a subset $\boldsymbol{\Theta}_0$ of $\boldsymbol{\Theta}$. For the multivariate normal situation with $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$ unspecified, $\boldsymbol{\Theta}_0 = \{\mu_1 = \mu_{10}, \mu_2 = \mu_{20}, \ldots, \mu_p = \mu_{p0}; \sigma_{11}, \ldots, \sigma_{1p}, \sigma_{22}, \ldots, \sigma_{2p}, \ldots, \sigma_{p-1,p}, \sigma_{pp}$ with $\boldsymbol{\Sigma}$ positive definite$\}$, so $\boldsymbol{\Theta}_0$ has dimension $\nu_0 = 0 + p(p+1)/2 = p(p+1)/2$.

A likelihood ratio test of H_0: $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ rejects H_0 in favor of H_1: $\boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$ if

$$\Lambda = \frac{\max\limits_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L(\boldsymbol{\theta})}{\max\limits_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta})} < c \tag{5-16}$$

Hotelling's T^2 and Likelihood Ratio Test Approximated by Chi-Square

A likelihood ratio test of $H_0: \boldsymbol{\theta} \in \Theta_0$ rejects H_0 in favor of $H_1: \boldsymbol{\theta} \notin \Theta_0$ if

$$\Lambda = \frac{\max\limits_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta})}{\max\limits_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta})} < c \tag{5-16}$$

where c is a suitably chosen constant. Intuitively, we reject H_0 if the maximum of the likelihood obtained by allowing $\boldsymbol{\theta}$ to vary over the set Θ_0 is much smaller than the maximum of the likelihood obtained by varying $\boldsymbol{\theta}$ over all values in Θ. When the maximum in the numerator of expression (5-16) is much smaller than the maximum in the denominator, Θ_0 does not contain plausible values for $\boldsymbol{\theta}$.

In each application of the likelihood ratio method, we must obtain the sampling distribution of the likelihood-ratio test statistic Λ. Then c can be selected to produce a test with a specified significance level α. However, when the sample size is large and certain regularity conditions are satisfied, the sampling distribution of $-2 \ln \Lambda$ is well approximated by a chi-square distribution. This attractive feature accounts, in part, for the popularity of likelihood ratio procedures.

Confidence Regions and Simultaneous Comparisons of Component Means

5.4 Confidence Regions and Simultaneous Comparisons of Component Means

To obtain our primary method for making inferences from a sample, we need to extend the concept of a univariate confidence interval to a multivariate *confidence region*. Let $\boldsymbol{\theta}$ be a vector of unknown population parameters and $\boldsymbol{\Theta}$ be the set of all possible values of $\boldsymbol{\theta}$. A confidence region is a region of likely $\boldsymbol{\theta}$ values. This region is determined by the data, and for the moment, we shall denote it by $R(\mathbf{X})$, where $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n]'$ is the data matrix.

The region $R(\mathbf{X})$ is said to be a $100(1-\alpha)\%$ *confidence region* if, before the sample is selected,

$$P[R(\mathbf{X}) \text{ will cover the true } \boldsymbol{\theta}] = 1 - \alpha \qquad (5\text{-}17)$$

This probability is calculated under the true, but unknown, value of $\boldsymbol{\theta}$.

The confidence region for the mean $\boldsymbol{\mu}$ of a p-dimensional normal population is available from (5-6). Before the sample is selected,

$$P\left[n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \leq \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right] = 1 - \alpha$$

For μ and∑, X-bar Will be Within The Probability or Area of 1-α

The confidence region for the mean $\boldsymbol{\mu}$ of a p-dimensional normal population is available from (5-6). Before the sample is selected,

$$P\left[n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \le \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right] = 1 - \alpha$$

whatever the values of the unknown $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. In words, $\overline{\mathbf{X}}$ will be within

$$[(n-1)pF_{p,n-p}(\alpha)/(n-p)]^{1/2}$$

of $\boldsymbol{\mu}$, with probability $1 - \alpha$, provided that distance is defined in terms of $n\mathbf{S}^{-1}$. For a particular sample, $\bar{\mathbf{x}}$ and $\mathbf{S}$ can be computed, and the inequality

$n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \le (n-1)pF_{p,n-p}(\alpha)/(n-p)$ will define a region $R(\mathbf{X})$ within the space of all possible parameter values. In this case, the region will be an ellipsoid centered at $\bar{\mathbf{x}}$. This ellipsoid is the $100(1-\alpha)\%$ confidence region for $\boldsymbol{\mu}$.

A 100% or 100(1-α) Confidence Region For the Mean of a p-Dimensional Normal Distribution Ellipsoid or Probability or Area of 1-α

A $100(1-\alpha)\%$ *confidence region* for the mean of a p-dimensional normal distribution is the ellipsoid determined by all $\boldsymbol{\mu}$ such that

$$n(\bar{\mathbf{x}}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \le \frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha) \qquad (5\text{-}18)$$

where $\bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j$, $\mathbf{S} = \frac{1}{(n-1)}\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'$ and $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are the sample observations.

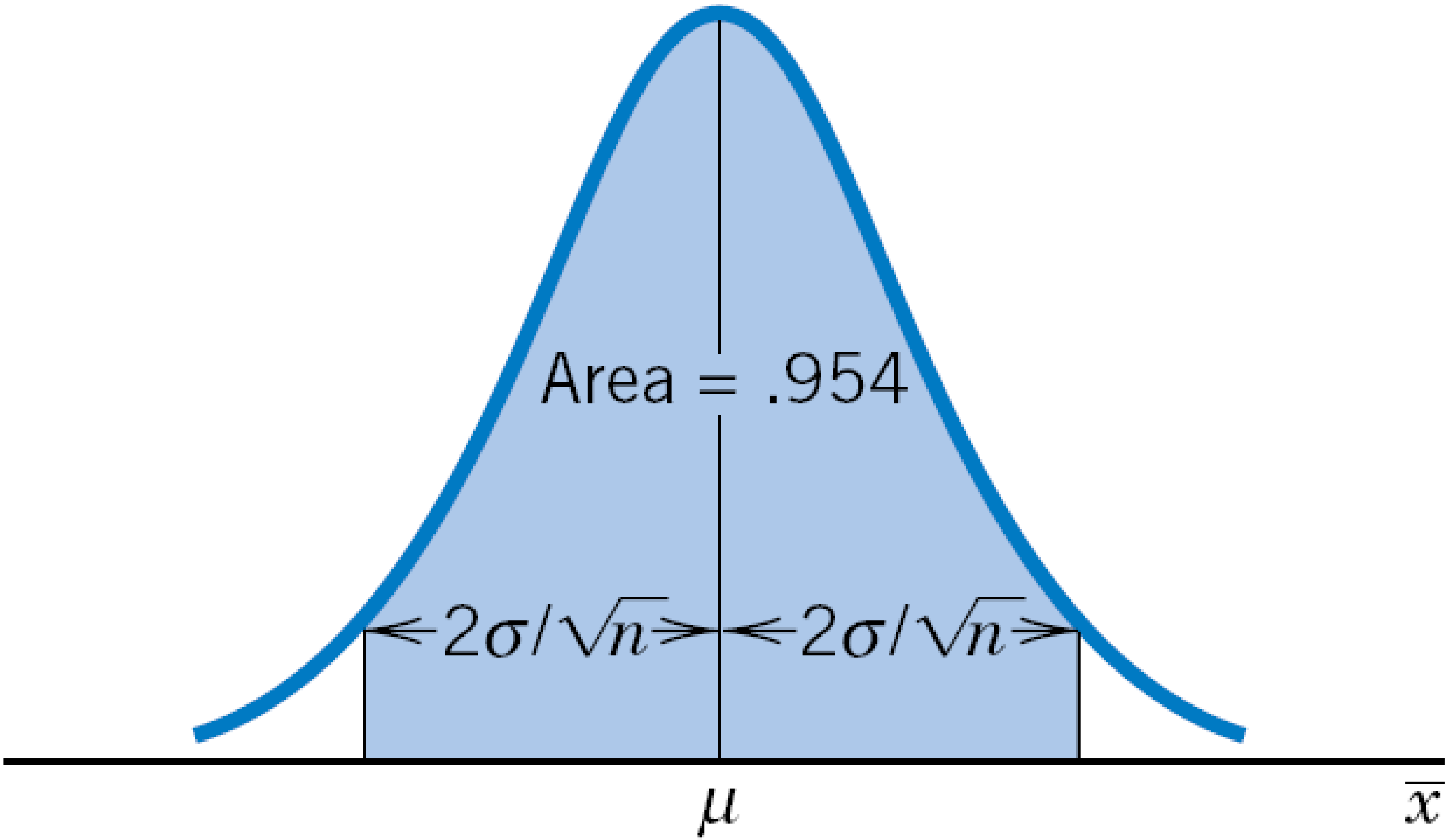


Figure 1 (p. 307)
Approximate normal distribution of $\overline{X}$.

Statistics, 7/E by Johnson and Bhattacharyya
Copyright © 2014 by John Wiley & Sons, Inc. All rights reserved.

Probability or Area Under the Curve Corresponds to 1-α or 100 (1-σ) %

Use of the probability .954, which corresponds to the multiplier 2 of the standard error, is by no means universal. The following notation will facilitate our writing of an expression for the $100\,(1-\alpha)$ % error margin, where $1-\alpha$ denotes the desired high probability such as .95 or .90.

Notation

$$z_{\alpha/2} = \text{Upper } \alpha/2 \text{ point of standard normal distribution}$$

That is, the area to the right of $z_{\alpha/2}$ is $\alpha/2$, and the area between $-z_{\alpha/2}$ and $z_{\alpha/2}$ is $1-\alpha$ (see Figure 2).

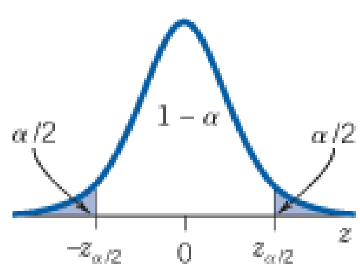


Figure 2 The notation $z_{\alpha/2}$.

Notation

$$z_{\alpha/2} = \text{Upper } \alpha/2 \text{ point of standard normal distribution}$$

That is, the area to the right of $z_{\alpha/2}$ is $\alpha/2$, and the area between $-z_{\alpha/2}$ and $z_{\alpha/2}$ is $1 - \alpha$ (see Figure 2).

Box on Page 308

Notation

Statistics, 7/E by Johnson and Bhattacharyya
Copyright © 2014 by John Wiley & Sons, Inc. All rights reserved.

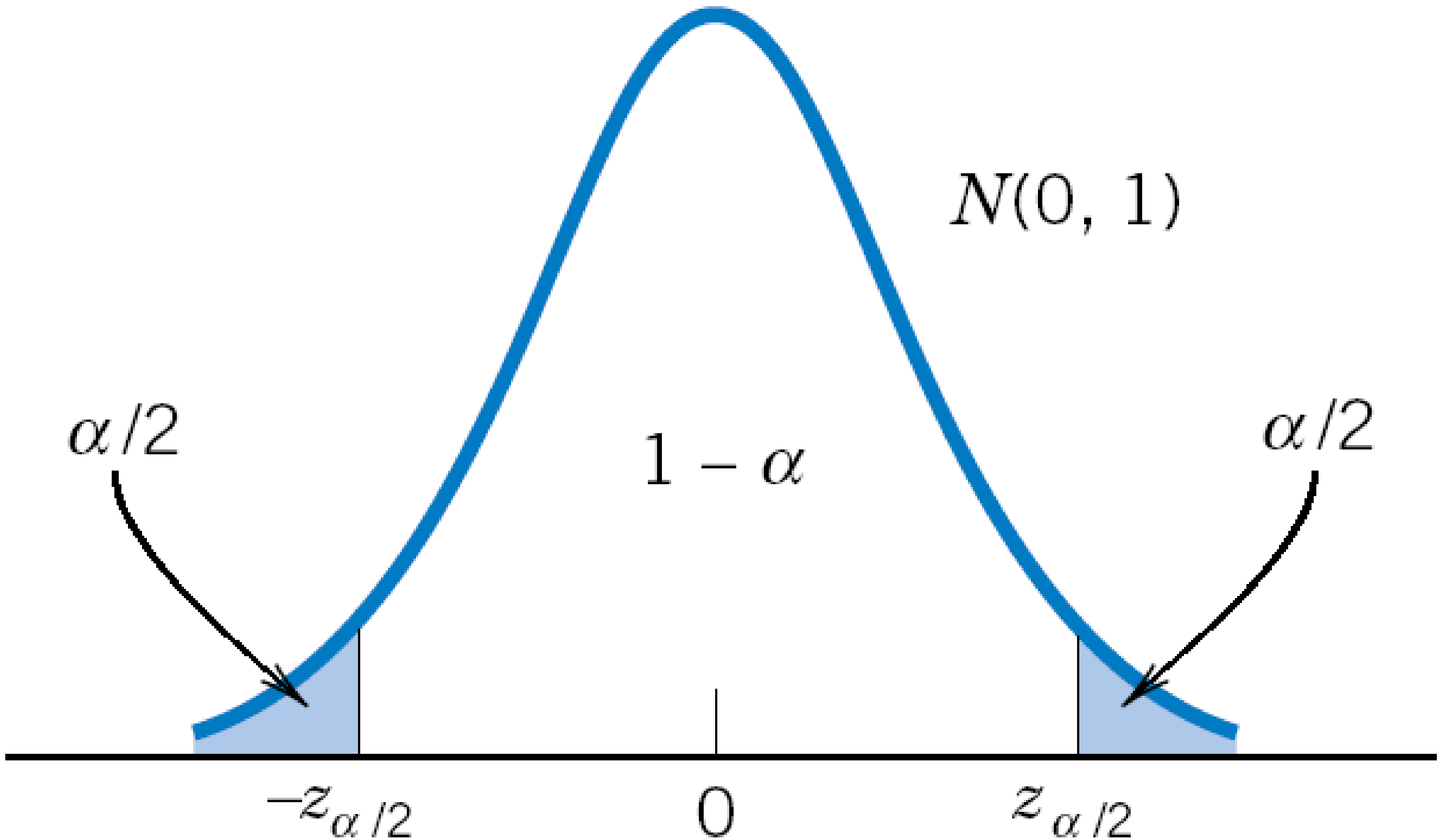


Figure 2 (p. 308)

The notation $z_{a/2}$.

Statistics, 7/E by Johnson and Bhattacharyya
Copyright © 2014 by John Wiley & Sons, Inc. All rights reserved.

From Z-Table: Values of $z_{\alpha/2}$ and Corresponding Z-Score

A few values of $z_{\alpha/2}$ obtained from the normal table appear in Table 2 for easy reference. To illustrate the notation, suppose we want to determine the 90% error margin. We then set $1-\alpha=.90$ so $\alpha/2=.05$ and, from Table 2, we have $z_{.05}=1.645$. Therefore, when estimating μ by $\overline{X}$, the 90% error margin is $1.645\ \sigma/\sqrt{n}$.

TABLE 2 Values of $z_{\alpha/2}$

$1-\alpha$	.80	.85	.90	.95	.99
$z_{\alpha/2}$	1.28	1.44	1.645	1.96	2.58

TABLE 2 Values of $z_{\alpha/2}$

$1 - \alpha$	.80	.85	.90	.95	.99
$z_{\alpha/2}$	1.28	1.44	1.645	1.96	2.58

Table 2 (p. 308)

Values of $z_{a/2}$

Copyright © 2014 by John Wiley & Sons, Inc. All rights reserved.

Standard Error as Point Estimator or Standard Deviation of μ

A minor difficulty remains in computing the standard error of $\overline{X}$. The expression involves the unknown population standard deviation σ, but we can estimate σ by the sample standard deviation.

$$S = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}}$$

When n is large, the effect of estimating the standard error $\sigma/\sqrt{n}$ by $S/\sqrt{n}$ can be neglected. We now summarize.

Point Estimation of the Mean

Parameter: Population mean μ.

Data: $X_1, \ldots, X_n$ (a random sample of size n)

Estimator: $\overline{X}$ (sample mean)

$$\text{S.E.}(\overline{X}) = \frac{\sigma}{\sqrt{n}} \qquad \text{Estimated S.E.}(\overline{X}) = \frac{S}{\sqrt{n}}$$

For large n, the $100(1 - \alpha)\%$ error margin is $z_{\alpha/2}\,\sigma/\sqrt{n}$.
(If σ is unknown, use S in place of σ.)

Box on Page 309

Point Estimation of the Mean

Statistics, 7/E by Johnson and Bhattacharyya
Copyright © 2014 by John Wiley & Sons, Inc. All rights reserved.

Confidence Interval Logic

Now, the relation

$$\mu - 1.96\frac{\sigma}{\sqrt{n}} < \overline{X} \quad \text{is the same as} \quad \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

and

$$\overline{X} < \mu + 1.96\frac{\sigma}{\sqrt{n}} \quad \text{is the same as} \quad \overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu$$

as we can see by transposing $1.96\,\sigma/\sqrt{n}$ from one side of an inequality to the other. Therefore, the event

$$\left[\mu - 1.96\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

Mean or µ Lies Between Standard Deviation Distance (Multiplied by Z-score) From x-bar, Both Negative and Positive

as we can see by transposing $1.96\,\sigma/\sqrt{n}$ from one side of an inequality to the other. Therefore, the event

$$\left[\mu - 1.96\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

is equivalent to

$$\left[\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

Thus, Can State With Certain Percent of Confidence or 95% Confidence Mean Lies Between Lower and Upper Bound

In essence, both events state that the difference $\overline{X} - \mu$ lies between $-1.96\ \sigma/\sqrt{n}$ and $1.96\ \sigma/\sqrt{n}$. Thus, the probability statement

$$P\left[\mu - 1.96\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu + 1.96\frac{\sigma}{\sqrt{n}}\right] = .95$$

can also be expressed as

$$P\left[\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right] = .95$$

Confidence Interval "Expression": Lower Bound, Upper Bound is 95% Confidence Interval for μ

This second form tells us that, before we sample, the random interval from $X - 1.96\ \sigma/\sqrt{n}$ to $X + 1.96\ \sigma/\sqrt{n}$ will include the unknown parameter μ with a probability of .95. Because σ is assumed to be known, both the upper and lower endpoints can be computed as soon as the sample data are available. Guided by the above reasonings, we say that the interval

$$\left(\overline{X} - 1.96\,\frac{\sigma}{\sqrt{n}},\quad \overline{X} + 1.96\,\frac{\sigma}{\sqrt{n}}\right)$$

or its realization $\left(\,\overline{x} - 1.96\ \sigma/\sqrt{n}\,, \overline{x} + 1.96\ \sigma/\sqrt{n}\right)$ is a **95% confidence interval for μ** when the population is normal and σ known.

Small p-Value Indicates Strong Evidence Again Ho or the Null Hypothesis

The P-value serves as a measure of the strength of evidence against H_0. A **small P-value** means that the null hypothesis is **strongly rejected** or the result is **highly statistically significant**.

Our illustrations of the basic concepts of hypothesis tests thus far focused on a problem where the alternative hypothesis is of the form $H_1: \mu < \mu_0$, called a left-sided alternative. If the alternative hypothesis in a problem states that the true μ is larger than its null hypothesis value of μ_0, we formulate the right-sided alternative $H_1: \mu > \mu_0$ and use a right-sided rejection region $R: Z \geq z_\alpha$.

The Steps for Testing Hypotheses

1. Formulate the null hypothesis H_0 and the alternative hypothesis H_1.
2. Test criterion: State the test statistic and the form of the rejection region.
3. With a specified α, determine the rejection region.
4. Calculate the test statistic from the data.
5. Draw a conclusion: State whether or not H_0 is rejected at the specified α and interpret the conclusion in the context of the problem. Also, it is a good statistical practice to calculate the P-value and strengthen the conclusion.

We illustrate the right-sided case and the main steps for conducting a statistical test as summarized above.

Calculate Generalized Squared Distance to Determine if In Confidence Region Based on the Axes of the Confidence Ellipsoid and Relative Lengths

To determine whether any $\boldsymbol{\mu}_0$ lies within the confidence region (is a plausible value for $\boldsymbol{\mu}$), we need to compute the generalized squared distance $n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$ and compare it with $[p(n-1)/(n-p)]F_{p,n-p}(\alpha)$. If the squared distance is larger than $[p(n-1)/(n-p)]F_{p,n-p}(\alpha)$, $\boldsymbol{\mu}_0$ is not in the confidence region. Since this is analogous to testing $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ [see (5-7)], we see that the confidence region of (5-18) consists of all $\boldsymbol{\mu}_0$ vectors for which the T^2-test would *not* reject H_0 in favor of H_1 at significance level α.

For $p \geq 4$, we cannot graph the joint confidence region for $\boldsymbol{\mu}$. However, we can calculate the axes of the confidence ellipsoid and their relative lengths. These are determined from the eigenvalues λ_i and eigenvectors $\mathbf{e}_i$ of $\mathbf{S}$. As in (4-7), the directions and lengths of the axes of

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \leq c^2 = \frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)$$

are determined by going

$$\sqrt{\lambda_i}\, c/\sqrt{n} = \sqrt{\lambda_i}\sqrt{p(n-1)F_{p,n-p}(\alpha)/n(n-p)}$$

Ratios of the Eigenvalues Will Help Identify Amount of Elongation, Thus Variance Along Pairs of Axes

are determined by going

$$\sqrt{\lambda_i}\, c/\sqrt{n} = \sqrt{\lambda_i}\sqrt{p(n-1)F_{p,n-p}(\alpha)/n(n-p)}$$

units along the eigenvectors $\mathbf{e}_i$. Beginning at the center $\bar{\mathbf{x}}$, the axes of the confidence ellipsoid are

$$\pm\sqrt{\lambda_i}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\,\mathbf{e}_i \qquad \text{where } \mathbf{S}\mathbf{e}_i = \lambda_i\mathbf{e}_i, \quad i = 1, 2, \ldots, p \qquad (5\text{-}19)$$

The ratios of the λ_i's will help identify relative amounts of elongation along pairs of axes.

Constructing a Confidence Ellipse for μ

Example 5.3 (Constructing a confidence ellipse for μ) Data for radiation from microwave ovens were introduced in Examples 4.10 and 4.17. Let

$$x_1 = \sqrt[4]{\text{measured radiation with door closed}}$$

and

$$x_2 = \sqrt[4]{\text{measured radiation with door open}}$$

Constructing a Confidence Ellipse for μ

For the $n = 42$ pairs of transformed observations, we find that

$$\bar{\mathbf{x}} = \begin{bmatrix} .564 \\ .603 \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} .0144 & .0117 \\ .0117 & .0146 \end{bmatrix},$$

$$\mathbf{S}^{-1} = \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix}$$

The eigenvalue and eigenvector pairs for $\mathbf{S}$ are

$$\lambda_1 = .026, \qquad \mathbf{e}_1' = [.704, \quad .710]$$

$$\lambda_2 = .002, \qquad \mathbf{e}_2' = [-.710, \quad .704]$$

The 95% confidence ellipse for $\boldsymbol{\mu}$ consists of all values (μ_1, μ_2) satisfying

$$42[.564 - \mu_1, \quad .603 - \mu_2] \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix} \begin{bmatrix} .564 - \mu_1 \\ .603 - \mu_2 \end{bmatrix} \le \frac{2(41)}{40} F_{2,40}(.05)$$

or, since $F_{2,40}(.05) = 3.23$,

$$42(203.018)(.564 - \mu_1)^2 + 42(200.228)(.603 - \mu_2)^2 - 84(163.391)(.564 - \mu_1)(.603 - \mu_2) \le 6.62$$

Joint Confidence Ellipsoid and Their Lengths Expressed as Eigenvectors

To see whether $\boldsymbol{\mu}' = [.562, .589]$ is in the confidence region, we compute

$$42(203.018)(.564 - .562)^2 + 42(200.228)(.603 - .589)^2 \\ - 84(163.391)(.564 - .562)(.603 - .589) = 1.30 \leq 6.62$$

We conclude that $\boldsymbol{\mu}' = [.562, .589]$ is in the region. Equivalently, a test of H_0: $\boldsymbol{\mu} = \begin{bmatrix} .562 \\ .589 \end{bmatrix}$ would not be rejected in favor of H_1: $\boldsymbol{\mu} \neq \begin{bmatrix} .562 \\ .589 \end{bmatrix}$ at the $\alpha = .05$ level of significance.

The joint confidence ellipsoid is plotted in Figure 5.1. The center is at $\bar{\mathbf{x}}' = [.564, .603]$, and the half-lengths of the major and minor axes are given by

$$\sqrt{\lambda_1}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)} = \sqrt{.026}\sqrt{\frac{2(41)}{42(40)}(3.23)} = .064$$

and

$$\sqrt{\lambda_2}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)} = \sqrt{.002}\sqrt{\frac{2(41)}{42(40)}(3.23)} = .018$$

respectively. The axes lie along $\mathbf{e}_1' = [.704, .710]$ and $\mathbf{e}_2' = [-.710, .704]$ when these vectors are plotted with $\bar{\mathbf{x}}$ as the origin. An indication of the elongation of the confidence ellipse is provided by the ratio of the lengths of the major and minor axes.

A 95 % Confidence Ellipse for μ Using Microwave Radiation Data

This ratio is

$$\frac{2\sqrt{\lambda_1}\sqrt{\dfrac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}}{2\sqrt{\lambda_2}\sqrt{\dfrac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}} = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} = \frac{.161}{.045} = 3.6$$

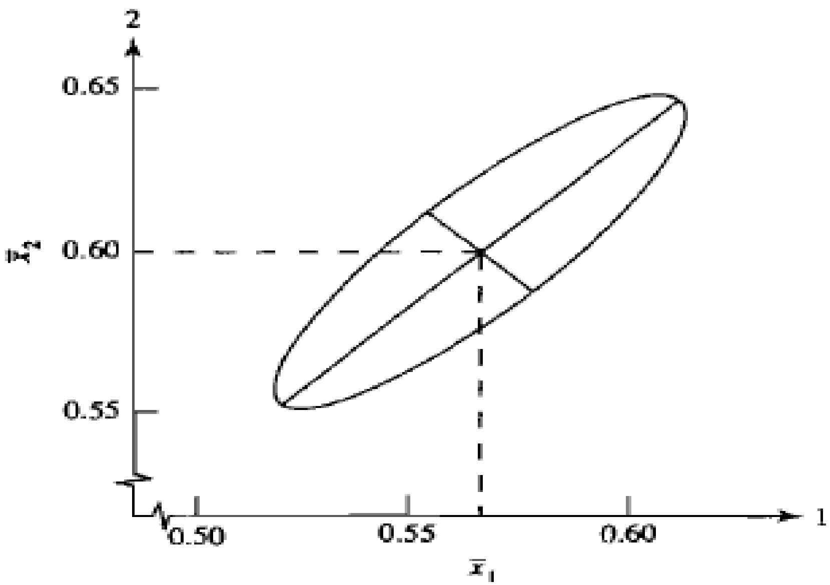


Figure 5.1 A 95% confidence ellipse for μ based on microwave-radiation data.

The length of the major axis is 3.6 times the length of the minor axis. ■

Simultaneous Confidence Statements Based on Variance

Simultaneous Confidence Statements

While the confidence region $n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \le c^2$, for c a constant, correctly assesses the joint knowledge concerning plausible values of $\boldsymbol{\mu}$, any summary of conclusions ordinarily includes confidence statements about the individual component means. In so doing, we adopt the attitude that all of the separate confidence statements should hold *simultaneously* with a specified high probability. It is the guarantee of a specified probability against *any* statement being incorrect that motivates the term *simultaneous confidence intervals*. We begin by considering simultaneous confidence statements which are intimately related to the joint confidence region based on the T^2-statistic.

Let $\mathbf{X}$ have an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution and form the linear combination

$$Z = a_1 X_1 + a_2 X_2 + \cdots + a_p X_p = \mathbf{a}'\mathbf{X}$$

From (2-43),

$$\mu_Z = E(Z) = \mathbf{a}'\boldsymbol{\mu}$$

and

$$\sigma_Z^2 = \mathrm{Var}(Z) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$$

Simultaneous Confidence Statements Based on Variance Using t-Test

Moreover, by Result 4.2, Z has an $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ distribution. If a random sample $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ from the $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population is available, a corresponding sample of Z's can be created by taking linear combinations. Thus,

$$Z_j = a_1 X_{j1} + a_2 X_{j2} + \cdots + a_p X_{jp} = \mathbf{a}'\mathbf{X}_j \qquad j = 1, 2, \ldots, n$$

The sample mean and variance of the observed values $z_1, z_2, \ldots, z_n$ are, by (3-36),

$$\bar{z} = \mathbf{a}'\bar{\mathbf{x}}$$
$$s_z^2 = \mathbf{a}'\mathbf{S}\mathbf{a}$$

where $\bar{\mathbf{x}}$ and $\mathbf{S}$ are the sample mean vector and covariance matrix of the $\mathbf{x}_j$'s, respectively.

Simultaneous confidence intervals can be developed from a consideration of confidence intervals for $\mathbf{a}'\boldsymbol{\mu}$ for various choices of $\mathbf{a}$. The argument proceeds as follows.

For $\mathbf{a}$ *fixed* and σ_Z^2 unknown, a $100(1 - \alpha)\%$ confidence interval for $\mu_Z = \mathbf{a}'\boldsymbol{\mu}$ is based on student's t-ratio

$$t = \frac{\bar{z} - \mu_Z}{s_z/\sqrt{n}} = \frac{\sqrt{n}\,(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu})}{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}} \qquad (5\text{-}20)$$

and leads to the statement

Confidence of All Statements Not 1-α

and leads to the statement

$$\bar{z} - t_{n-1}(\alpha/2)\frac{s_z}{\sqrt{n}} \leq \mu_Z \leq \bar{z} + t_{n-1}(\alpha/2)\frac{s_z}{\sqrt{n}}$$

or

$$\mathbf{a}'\bar{\mathbf{x}} - t_{n-1}(\alpha/2)\frac{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}}{\sqrt{n}} \leq \mathbf{a}'\boldsymbol{\mu} \leq \mathbf{a}'\bar{\mathbf{x}} + t_{n-1}(\alpha/2)\frac{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}}{\sqrt{n}} \tag{5-21}$$

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

Inequality (5-21) can be interpreted as a statement about the components of the mean vector $\boldsymbol{\mu}$. For example, with $\mathbf{a}' = [1, 0, \ldots, 0]$, $\mathbf{a}'\boldsymbol{\mu} = \mu_1$, and (5-21) becomes the usual confidence interval for a normal population mean. (Note, in this case, that $\mathbf{a}'\mathbf{S}\mathbf{a} = s_{11}$.) Clearly, we could make several confidence statements about the components of $\boldsymbol{\mu}$, each with associated confidence coefficient $1 - \alpha$, by choosing different coefficient vectors $\mathbf{a}$. However, the confidence associated with all of the statements taken together is *not* $1 - \alpha$.

Large Simultaneous Confidence Intervals Are Wider Than 1-α

Intuitively, it would be desirable to associate a "collective" confidence coefficient of $1 - \alpha$ with the confidence intervals that can be generated by all choices of $\mathbf{a}$. However, a price must be paid for the convenience of a large simultaneous confidence coefficient: intervals that are wider (less precise) than the interval of (5-21) for a specific choice of $\mathbf{a}$.

Given a data set $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ and a particular $\mathbf{a}$, the confidence interval in (5-21) is that set of $\mathbf{a}'\boldsymbol{\mu}$ values for which

$$|t| = \left| \frac{\sqrt{n}\,(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu})}{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}} \right| \le t_{n-1}(\alpha/2)$$

or, equivalently,

$$t^2 = \frac{n(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu})^2}{\mathbf{a}'\mathbf{S}\mathbf{a}} = \frac{n(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}))^2}{\mathbf{a}'\mathbf{S}\mathbf{a}} \le t_{n-1}^2(\alpha/2) \qquad (5\text{-}22)$$

A simultaneous confidence region is given by the set of $\mathbf{a}'\boldsymbol{\mu}$ values such that t^2 is relatively small for *all* choices of $\mathbf{a}$. It seems reasonable to expect that the constant $t_{n-1}^2(\alpha/2)$ in (5-22) will be replaced by a larger value, c^2, when statements are developed for many choices of $\mathbf{a}$.

Large Simultaneous Confidence Intervals Are Wider Than 1-α Due to Volume of the Ellipsoid

Considering the values of $\mathbf{a}$ for which $t^2 \le c^2$, we are naturally led to the determination of

$$\max_{\mathbf{a}} t^2 = \max_{\mathbf{a}} \frac{n(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}))^2}{\mathbf{a}'\mathbf{S}\mathbf{a}}$$

Using the maximization lemma (2-50) with $\mathbf{x} = \mathbf{a}$, $\mathbf{d} = (\bar{\mathbf{x}} - \boldsymbol{\mu})$, and $\mathbf{B} = \mathbf{S}$, we get

$$\max_{\mathbf{a}} \frac{n(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}))^2}{\mathbf{a}'\mathbf{S}\mathbf{a}} = n\left[\max_{\mathbf{a}} \frac{(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}))^2}{\mathbf{a}'\mathbf{S}\mathbf{a}}\right] = n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) = T^2 \quad (5\text{-}23)$$

with the maximum occurring for $\mathbf{a}$ proportional to $\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$.

Result 5.3. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population with $\boldsymbol{\Sigma}$ positive definite. Then, simultaneously for all $\mathbf{a}$, the interval

$$\left(\mathbf{a}'\overline{\mathbf{X}} - \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)\mathbf{a}'\mathbf{S}\mathbf{a}}, \quad \mathbf{a}'\overline{\mathbf{X}} + \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)\mathbf{a}'\mathbf{S}\mathbf{a}}\right)$$

will contain $\mathbf{a}'\boldsymbol{\mu}$ with probability $1 - \alpha$.

Proof:Large Simultaneous Confidence Intervals Are Wider Than 1-α Due to Volume of the Ellipsoid

Proof. From (5-23),

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \le c^2 \quad \text{implies} \quad \frac{n(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu})^2}{\mathbf{a}'\mathbf{S}\mathbf{a}} \le c^2$$

for every $\mathbf{a}$, or

$$\mathbf{a}'\bar{\mathbf{x}} - c\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}} \le \mathbf{a}'\boldsymbol{\mu} \le \mathbf{a}'\bar{\mathbf{x}} + c\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$$

for every $\mathbf{a}$. Choosing $c^2 = p(n-1)F_{p,n-p}(\alpha)/(n-p)$ [see (5-6)] gives intervals that will contain $\mathbf{a}'\boldsymbol{\mu}$ for all $\mathbf{a}$, with probability $1 - \alpha = P[T^2 \le c^2]$. ■

Simultaneous Intervals Referred To As T^2 Intervals

It is convenient to refer to the simultaneous intervals of Result 5.3 as T^2-*intervals*, since the coverage probability is determined by the distribution of T^2. The successive choices $\mathbf{a}' = [1, 0, \dots, 0]$, $\mathbf{a}' = [0, 1, \dots, 0]$, and so on through $\mathbf{a}' = [0, 0, \dots, 1]$ for the T^2-intervals allow us to conclude that

$$\begin{aligned}
\bar{x}_1 - \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{11}}{n}} \le \mu_1 \le \bar{x}_1 + \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{11}}{n}} \\
\bar{x}_2 - \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{22}}{n}} \le \mu_2 \le \bar{x}_2 + \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{22}}{n}} \\
\vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots \\
\bar{x}_p - \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{pp}}{n}} \le \mu_p \le \bar{x}_p + \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{pp}}{n}}
\end{aligned} \tag{5-24}$$

all hold simultaneously with confidence coefficient $1 - \alpha$. Note that, without modifying the coefficient $1 - \alpha$, we can make statements about the differences $\mu_i - \mu_k$ corresponding to $\mathbf{a}' = [0, \dots, 0, \ a_i, 0, \dots, 0, \ a_k, 0, \dots, 0]$, where $a_i = 1$ and

Simultaneous Intervals Referred To As T^2 Intervals Ideal for Data Patterns

$a_k = -1$. In this case $\mathbf{a}'\mathbf{S}\mathbf{a} = s_{ii} - 2s_{ik} + s_{kk}$, and we have the statement

$$\bar{x}_i - \bar{x}_k - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii} - 2s_{ik} + s_{kk}}{n}} \le \mu_i - \mu_k$$
$$\le \bar{x}_i - \bar{x}_k + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii} - 2s_{ik} + s_{kk}}{n}} \qquad (5\text{-}25)$$

The simultaneous T^2 confidence intervals are ideal for "data snooping." The confidence coefficient $1 - \alpha$ remains unchanged for any choice of $\mathbf{a}$, so linear combinations of the components μ_i that merit inspection *based upon an examination of the data* can be estimated.

Simultaneous Intervals T^2 Intervals Of A Mean Vector Are Projections of the Confidence Ellipsoid on the Component Axes

In addition, according to the results in Supplement 5A, we can include the statements about (μ_i, μ_k) belonging to the sample mean-centered ellipses

$$n[\bar{x}_i - \mu_i, \quad \bar{x}_k - \mu_k] \begin{bmatrix} s_{ii} & s_{ik} \\ s_{ik} & s_{kk} \end{bmatrix}^{-1} \begin{bmatrix} \bar{x}_i - \mu_i \\ \bar{x}_k - \mu_k \end{bmatrix} \leq \frac{p(n-1)}{n-p} F_{p,n-p}(\alpha) \qquad (5\text{-}26)$$

and still maintain the confidence coefficient $(1 - \alpha)$ for the whole set of statements.

The simultaneous T^2 confidence intervals for the individual components of a mean vector are just the shadows, or projections, of the confidence ellipsoid on the component axes. This connection between the shadows of the ellipsoid and the simultaneous confidence intervals given by (5-24) is illustrated in the next example.

Example: Simultaneous Intervals T^2 Intervals Of A Mean Vector Are Projections of the Confidence Ellipsoid

Example 5.4 (Simultaneous confidence intervals as shadows of the confidence ellipsoid) In Example 5.3, we obtained the 95% confidence ellipse for the means of the fourth roots of the door-closed and door-open microwave radiation measurements. The 95% simultaneous T^2 intervals for the two component means are, from (5-24),

$$\left(\bar{x}_1 - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(.05)}\sqrt{\frac{s_{11}}{n}},\quad \bar{x}_1 + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(.05)}\sqrt{\frac{s_{11}}{n}}\right)$$

$$= \left(.564 - \sqrt{\frac{2(41)}{40}3.23}\sqrt{\frac{.0144}{42}},\quad .564 + \sqrt{\frac{2(41)}{40}3.23}\sqrt{\frac{.0144}{42}}\right) \quad \text{or} \quad (.516,\quad .612)$$

$$\left(\bar{x}_2 - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(.05)}\sqrt{\frac{s_{22}}{n}},\quad \bar{x}_2 + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(.05)}\sqrt{\frac{s_{22}}{n}}\right)$$

$$= \left(.603 - \sqrt{\frac{2(41)}{40}3.23}\sqrt{\frac{.0146}{42}},\quad .603 + \sqrt{\frac{2(41)}{40}3.23}\sqrt{\frac{.0146}{42}}\right) \quad \text{or} \quad (.555,\quad .651)$$

In Figure 5.2, we have redrawn the 95% confidence ellipse from Example 5.3. The 95% simultaneous intervals are shown as shadows, or projections, of this ellipse on the axes of the component means. ■

Example: Simultaneous Intervals T^2 Intervals Of A Mean Vector Are Projections of the Confidence Ellipsoid

Example 5.5 (Constructing simultaneous confidence intervals and ellipses) The scores obtained by $n = 87$ college students on the College Level Examination Program (CLEP) subtest X_1 and the College Qualification Test (CQT) subtests X_2 and X_3 are given in Table 5.2 on page 228 for X_1 = social science and history, X_2 = verbal, and X_3 = science. These data give

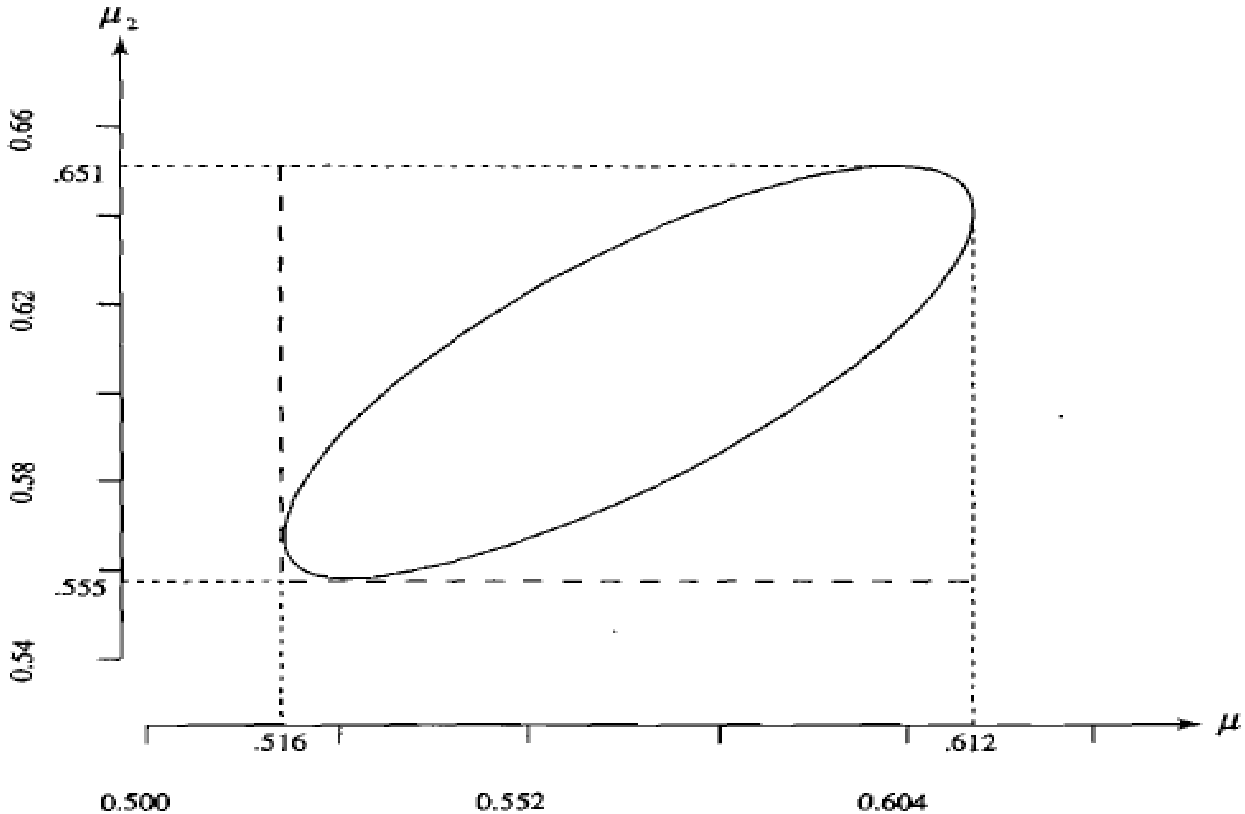


Figure 5.2 Simultaneous T^2-intervals for the component means as shadows of the confidence ellipse on the axes—microwave radiation data.

Simultaneous Intervals T^2 Intervals Of A Mean Vector As Projections of the Confidence Ellipsoid

$$\bar{\mathbf{x}} = \begin{bmatrix} 526.29 \\ 54.69 \\ 25.13 \end{bmatrix} \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} 5808.06 & 597.84 & 222.03 \\ 597.84 & 126.05 & 23.39 \\ 222.03 & 23.39 & 23.11 \end{bmatrix}$$

Let us compute the 95% simultaneous confidence intervals for μ_1, μ_2, and μ: We have

$$\frac{p(n-1)}{n-p} F_{p,n-p}(\alpha) = \frac{3(87-1)}{(87-3)} F_{3,84}(.05) = \frac{3(86)}{84}(2.7) = 8.29$$

and we obtain the simultaneous confidence statements [see (5-24)]

$$526.29 - \sqrt{8.29}\sqrt{\frac{5808.06}{87}} \leq \mu_1 \leq 526.29 + \sqrt{8.29}\sqrt{\frac{5808.06}{87}}$$

or

$$503.06 \leq \mu_1 \leq 550.12$$

$$54.69 - \sqrt{8.29}\sqrt{\frac{126.05}{87}} \leq \mu_2 \leq 54.69 + \sqrt{8.29}\sqrt{\frac{126.05}{87}}$$

or

$$51.22 \leq \mu_2 \leq 58.16$$

$$25.13 - \sqrt{8.29}\sqrt{\frac{23.11}{87}} \leq \mu_3 \leq 25.13 + \sqrt{8.29}\sqrt{\frac{23.11}{87}}$$

College Test Data

Table 5.2 College Test Data

Individual	X_1 (Social science and history)	X_2 (Verbal)	X_3 (Science)	Individual	X_1 (Social science and history)	X_2 (Verbal)	X_3 (Science)
1	468	41	26	45	494	41	24
2	428	39	26	46	541	47	25
3	514	53	21	47	362	36	17
4	547	67	33	48	408	28	17
5	614	61	27	49	594	68	23
6	501	67	29	50	501	25	26
7	421	46	22	51	687	75	33
8	527	50	23	52	633	52	31
9	527	55	19	53	647	67	29
10	620	72	32	54	647	65	34
11	587	63	31	55	614	59	25
12	541	59	19	56	633	65	28
13	561	53	26	57	448	55	24
14	468	62	20	58	408	51	19
15	614	65	28	59	441	35	22
16	527	48	21	60	435	60	20
17	507	32	27	61	501	54	21
18	580	64	21	62	507	42	24
19	507	59	21	63	620	71	36
20	521	54	23	64	415	52	20
21	574	52	25	65	554	69	30
22	587	64	31	66	348	28	18
23	488	51	27	67	468	49	25
24	488	62	18	68	507	54	26
25	587	56	26	69	527	47	31
26	421	38	16	70	527	47	26
27	481	52	26	71	435	50	28
28	428	40	19	72	660	70	25
29	640	65	25	73	733	73	33
30	574	61	28	74	507	45	28
31	547	64	27	75	527	62	29
32	580	64	28	76	428	37	19

College Test Data – Cont.

32	580	64	28	76	428	37	19
33	494	53	26	77	481	48	23
34	554	51	21	78	507	61	19
35	647	58	23	79	527	66	23
36	507	65	23	80	488	41	28
37	454	52	28	81	607	69	28
38	427	57	21	82	561	59	34
39	521	66	26	83	614	70	23
40	468	57	14	84	527	49	30
41	587	55	30	85	474	41	16
42	507	61	31	86	441	47	26
43	574	54	31	87	607	67	32
44	507	53	23				

Source: Data courtesy of Richard W. Johnson.

Simultaneous Intervals T^2 Intervals Wider Than Univariate Intervals Because All Most Hold 95% Confidence

or

$$23.65 \leq \mu_3 \leq 26.61$$

With the possible exception of the verbal scores, the marginal Q–Q plots and two-dimensional scatter plots do not reveal any serious departures from normality for the college qualification test data. (See Exercise 5.18.) Moreover, the sample size is large enough to justify the methodology, even though the data are not quite normally distributed. (See Section 5.5.)

The simultaneous T^2-intervals above are wider than univariate intervals because all three must hold with 95% confidence. They may also be wider than necessary, because, with the same confidence, we can make statements about differences.

For instance, with $\mathbf{a}' = [0, 1, -1]$, the interval for $\mu_2 - \mu_3$ has endpoints

$$(\bar{x}_2 - \bar{x}_3) \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(.05)} \sqrt{\frac{s_{22} + s_{33} - 2s_{23}}{n}}$$

$$= (54.69 - 25.13) \pm \sqrt{8.29} \sqrt{\frac{126.05 + 23.11 - 2(23.39)}{87}} = 29.56 \pm 3.12$$

so (26.44, 32.68) is a 95% confidence interval for $\mu_2 - \mu_3$. Simultaneous intervals can also be constructed for the other differences.

Simultaneous Intervals T^2 Intervals Of A Mean Vector Are Projections of the Confidence Ellipsoid

Finally, we can construct confidence ellipses for pairs of means, and the same 95% confidence holds. For example, for the pair (μ_2, μ_3), we have

$$
\begin{aligned}
&87[54.69 - \mu_2, \quad 25.13 - \mu_3] \begin{bmatrix} 126.05 & 23.39 \\ 23.39 & 23.11 \end{bmatrix}^{-1} \begin{bmatrix} 54.69 - \mu_2 \\ 25.13 - \mu_3 \end{bmatrix} \\
&\quad = 0.849(54.69 - \mu_2)^2 + 4.633(25.13 - \mu_3)^2 \\
&\qquad\qquad - 2 \times 0.859(54.69 - \mu_2)(25.13 - \mu_3) \le 8.29
\end{aligned}
$$

This ellipse is shown in Figure 5.3 on page 230, along with the 95% confidence ellipses for the other two pairs of means. The projections or shadows of these ellipses on the axes are also indicated, and these projections are the T^2-intervals. ■

Comparison Simultaneous Confidence Intervals – One At A Time

A Comparison of Simultaneous Confidence Intervals with One-at-a-Time Intervals

An alternative approach to the construction of confidence intervals is to consider the components μ_i one at a time, as suggested by (5-21) with $\mathbf{a}' = [0, \ldots, 0, a_i, 0, \ldots, 0]$ where $a_i = 1$. This approach ignores the covariance structure of the p variables and leads to the intervals

$$\begin{aligned}
\bar{x}_1 - t_{n-1}(\alpha/2)\sqrt{\frac{s_{11}}{n}} &\le \mu_1 \le \bar{x}_1 + t_{n-1}(\alpha/2)\sqrt{\frac{s_{11}}{n}} \\
\bar{x}_2 - t_{n-1}(\alpha/2)\sqrt{\frac{s_{22}}{n}} &\le \mu_2 \le \bar{x}_2 + t_{n-1}(\alpha/2)\sqrt{\frac{s_{22}}{n}} \\
&\;\;\vdots \\
\bar{x}_p - t_{n-1}(\alpha/2)\sqrt{\frac{s_{pp}}{n}} &\le \mu_p \le \bar{x}_p + t_{n-1}(\alpha/2)\sqrt{\frac{s_{pp}}{n}}
\end{aligned} \tag{5-27}$$

Simultaneous Intervals T^2 Intervals Of A Mean Vector Are Projections of the Confidence Ellipsoid – Shown Individually College Test Data

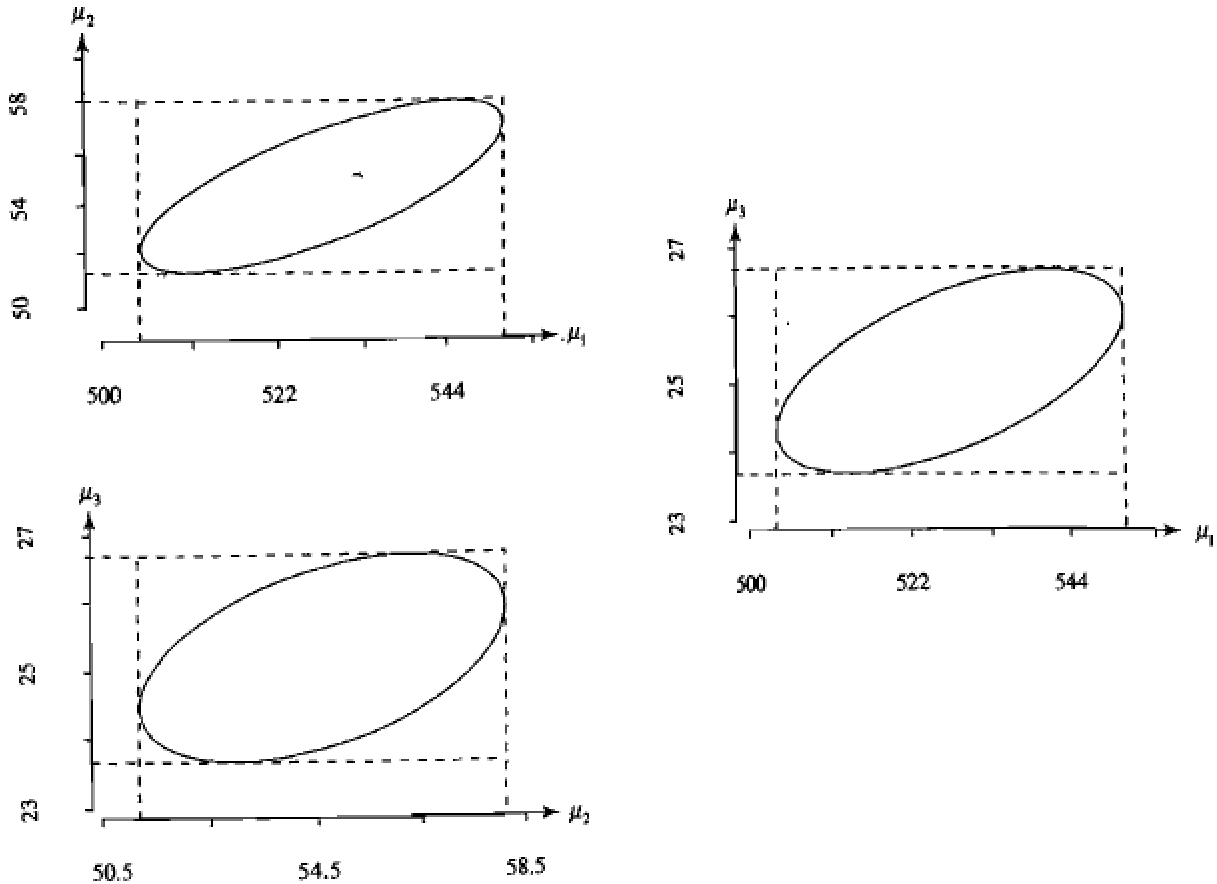


Figure 5.3 95% confidence ellipses for pairs of means and the simultaneous T^2-intervals—college test data.

If 1-α = .95, and p = 6, this probability is $(.95)^6 = .74$

Although prior to sampling, the ith interval has probability $1 - \alpha$ of covering μ_i, we do not know what to assert, in general, about the probability of *all* intervals containing their respective μ_i's. As we have pointed out, this probability is not $1 - \alpha$.

To shed some light on the problem, consider the special case where the observations have a joint normal distribution and

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}$$

Since the observations on the first variable are independent of those on the second variable, and so on, the product rule for independent events can be applied. Before the sample is selected,

$$P[\text{all } t\text{-intervals in (5-27) contain the } \mu_i\text{'s}] = (1-\alpha)(1-\alpha)\cdots(1-\alpha) = (1-\alpha)^p$$

If $1 - \alpha = .95$ and $p = 6$, this probability is $(.95)^6 = .74$.

Width of T^2 intervals Relative to t-intervals, Increases for Fixed n as p Increases and Decreases for Fixed p as n increases

To guarantee a probability of $1 - \alpha$ that all of the statements about the component means hold simultaneously, the individual intervals must be wider than the separate t-intervals; just how much wider depends on both p and n, as well as on $1 - \alpha$.

For $1 - \alpha = .95$, $n = 15$, and $p = 4$, the multipliers of $\sqrt{s_{ii}/n}$ in (5-24) and (5-27) are

$$\sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(.05)} = \sqrt{\frac{4(14)}{11}(3.36)} = 4.14$$

and $t_{n-1}(.025) = 2.145$, respectively. Consequently, in this case the simultaneous intervals are $100(4.14 - 2.145)/2.145 = 93\%$ wider than those derived from the one-at-a-time t method.

Table 5.3 gives some critical distance multipliers for one-at-a-time t-intervals computed according to (5-21), as well as the corresponding simultaneous T^2-intervals. In general, the width of the T^2-intervals, relative to the t-intervals, increases for fixed n as p increases and decreases for fixed p as n increases.

Critical Distance Multipliers for One-At-A-Time t-Intervals and T^2-Intervals

Table 5.3 gives some critical distance multipliers for one-at-a-time t-intervals computed according to (5-21), as well as the corresponding simultaneous T^2-intervals. In general, the width of the T^2-intervals, relative to the t-intervals, increases for fixed n as p increases and decreases for fixed p as n increases.

Table 5.3 Critical Distance Multipliers for One-at-a-Time t- Intervals and T^2-Intervals for Selected n and p $(1 - \alpha = .95)$

		$\sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(.05)}$	
n	$t_{n-1}(.025)$	$p = 4$	$p = 10$
15	2.145	4.14	11.52
25	2.064	3.60	6.39
50	2.010	3.31	5.05
100	1.970	3.19	4.61
∞	1.960	3.08	4.28

Critical Distance Multipliers for One-At-A-Time t-Intervals and T^2-Intervals: Use Both Critically

Table 5.3 Critical Distance Multipliers for One-at-a-Time t- Intervals and T^2-Intervals for Selected n and p $(1 - \alpha = .95)$

		$\sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(.05)}$	
n	$t_{n-1}(.025)$	$p = 4$	$p = 10$
15	2.145	4.14	11.52
25	2.064	3.60	6.39
50	2.010	3.31	5.05
100	1.970	3.19	4.61
∞	1.960	3.08	4.28

The comparison implied by Table 5.3 is a bit unfair, since the confidence level associated with any collection of T^2-intervals, for fixed n and p, is .95, and the overall confidence associated with a collection of individual t intervals, for the same n, can, as we have seen, be much less than .95. The one-at-a-time t intervals are too short to maintain an overall confidence level for separate statements about, say, all p means. Nevertheless, we sometimes look at them as the best possible information concerning a mean, if this is the only inference to be made. Moreover, if the one-at-a-time intervals are calculated only when the T^2-test rejects the null hypothesis, some researchers think they may more accurately represent the information about the means than the T^2-intervals do.

When T^2-Intervals Cover More than 95% Area, Second Approach To Multiple Comparisons Bonferroni Method

The T^2-intervals are too wide if they are applied only to the p component means. To see why, consider the confidence ellipse and the simultaneous intervals shown in Figure 5.2. If μ_1 lies in its T^2-interval and μ_2 lies in its T^2-interval, then (μ_1, μ_2) lies in the rectangle formed by these two intervals. This rectangle contains the confidence ellipse and more. The confidence ellipse is smaller but has probability .95 of covering the mean vector $\boldsymbol{\mu}$ with its component means μ_1 and μ_2. Consequently, the probability of covering the two individual means μ_1 and μ_2 will be larger than .95 for the rectangle formed by the T^2-intervals. This result leads us to consider a second approach to making multiple comparisons known as the Bonferroni method.

Bonferroni Method of Multiple Comparisons

The Bonferroni Method of Multiple Comparisons

Often, attention is restricted to a small number of individual confidence statements. In these situations it is possible to do better than the simultaneous intervals of Result 5.3. If the number m of specified component means μ_i or linear combinations $\mathbf{a}'\boldsymbol{\mu} = a_1\mu_1 + a_2\mu_2 + \cdots + a_p\mu_p$ is small, simultaneous confidence intervals can be developed that are shorter (more precise) than the simultaneous T^2-intervals. The alternative method for multiple comparisons is called the *Bonferroni method*, because it is developed from a probability inequality carrying that name.

Suppose that, prior to the collection of data, confidence statements about m linear combinations $\mathbf{a}_1'\boldsymbol{\mu}, \mathbf{a}_2'\boldsymbol{\mu}, \ldots, \mathbf{a}_m'\boldsymbol{\mu}$ are required. Let C_i denote a confidence statement about the value of $\mathbf{a}_i'\boldsymbol{\mu}$ with $P[C_i \text{ true}] = 1 - \alpha_i$, $i = 1, 2, \ldots, m$. Now (see Exercise 5.6),

$$
\begin{aligned}
P[\text{all } C_i \text{ true}] &= 1 - P[\text{at least one } C_i \text{ false}] \\
&\geq 1 - \sum_{i=1}^{m} P(C_i \text{ false}) = 1 - \sum_{i=1}^{m} (1 - \mathrm{P}(C_i \text{ true})) \\
&= 1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_m) \qquad (5\text{-}28)
\end{aligned}
$$

Bonferroni Method: Shorter Confidence Intervals By Controlling Error Rate

Inequality (5-28), a special case of the Bonferroni inequality, allows an investigator to control the overall error rate $\alpha_1 + \alpha_2 + \cdots + \alpha_m$, regardless of the correlation structure behind the confidence statements. There is also the flexibility of controlling the error rate for a group of important statements and balancing it by another choice for the less important statements.

Let us develop simultaneous interval estimates for the restricted set consisting of the components μ_i of $\boldsymbol{\mu}$. Lacking information on the relative importance of these components, we consider the individual t-intervals

$$\bar{x}_i \pm t_{n-1}\left(\frac{\alpha_i}{2}\right)\sqrt{\frac{s_{ii}}{n}} \qquad i = 1, 2, \ldots, m$$

with $\alpha_i = \alpha/m$. Since $P[\bar{X}_i \pm t_{n-1}(\alpha/2m)\sqrt{s_{ii}/n}$ contains $\mu_i] = 1 - \alpha/m$, $i = 1, 2, \ldots, m$, we have, from (5-28),

$$P\left[\bar{X}_i \pm t_{n-1}\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}} \text{ contains } \mu_i, \text{ all } i\right] \geq 1 - \underbrace{\left(\frac{\alpha}{m} + \frac{\alpha}{m} + \cdots + \frac{\alpha}{m}\right)}_{m \text{ terms}}$$

$$= 1 - \alpha$$

Bonferroni Method: Shorter Confidence Intervals By Controlling Error Rate: Same Structure as Hotellings- T^2

Therefore, with an overall confidence level greater than or equal to $1 - \alpha$, we can make the following $m = p$ statements:

$$\begin{aligned}
\bar{x}_1 - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}} &\le \mu_1 \le \bar{x}_1 + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}} \\
\bar{x}_2 - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{22}}{n}} &\le \mu_2 \le \bar{x}_2 + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{22}}{n}} \\
\vdots \qquad\qquad &\quad \vdots \qquad\qquad \vdots \\
\bar{x}_p - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}} &\le \mu_p \le \bar{x}_p + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}}
\end{aligned} \tag{5-29}$$

The statements in (5-29) can be compared with those in (5-24). The percentage point $t_{n-1}(\alpha/2p)$ replaces $\sqrt{(n-1)pF_{p,n-p}(\alpha)/(n-p)}$, but otherwise the intervals are of the same structure.