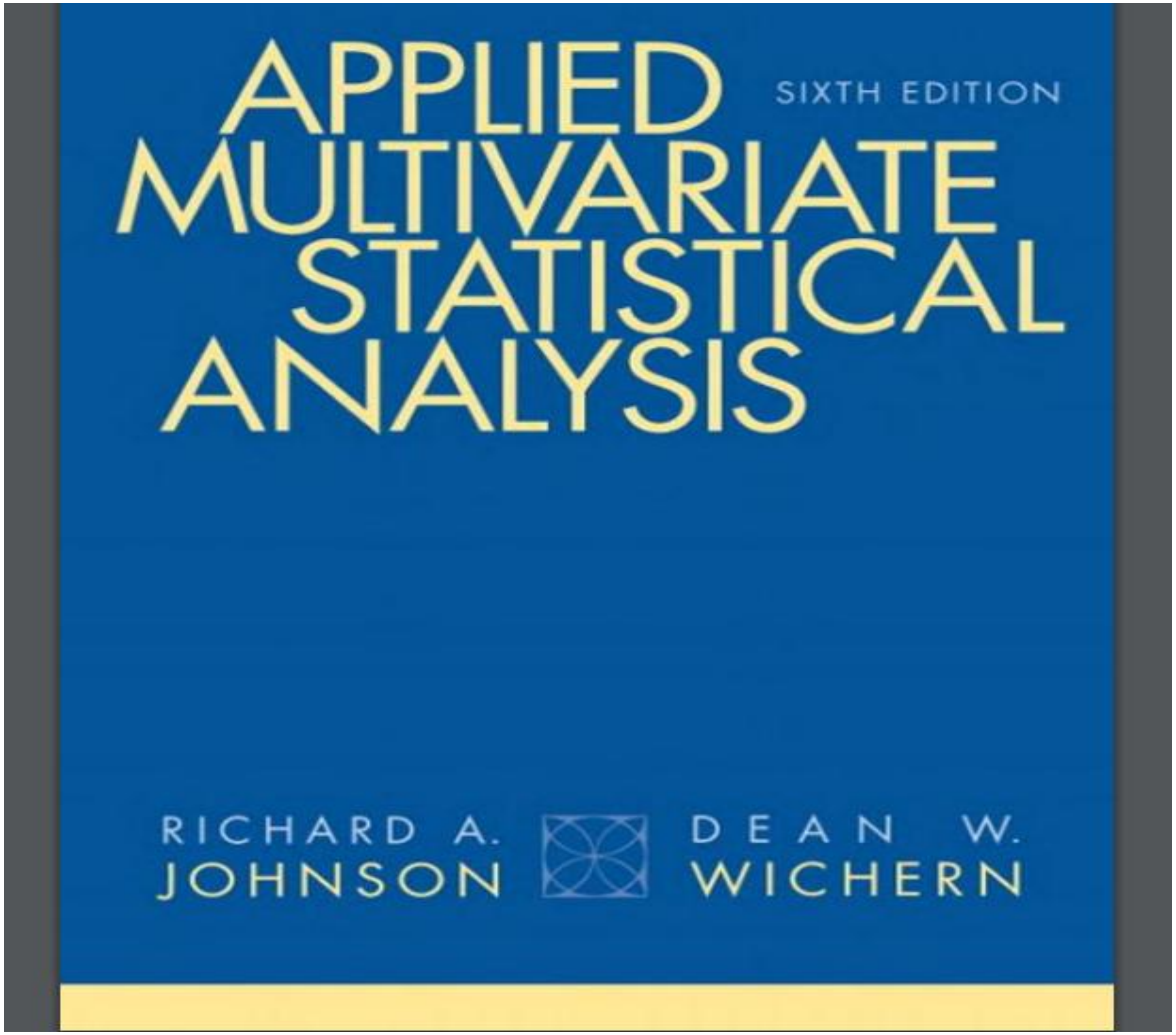
Advanced Multivariate Methods



Comparison of Several Means: ANOVA and MANOVA

COMPARISONS OF SEVERAL MULTIVARIATE MEANS

6.1 Introduction

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

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

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

Paired Comparisons and Repeated Measures Design

6.2 Paired Comparisons and a Repeated Measures Design

Paired Comparisons

Measurements are often recorded under different sets of experimental conditions to see whether the responses differ significantly over these sets. For example, the efficacy of a new drug or of a saturation advertising campaign may be determined by comparing measurements before the "treatment" (drug or advertising) with those

after the treatment. In other situations, *two or more* treatments can be administered to the same or similar experimental units, and responses can be compared to assess the effects of the treatments.

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

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

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

should reflect only the differential effects of the treatments.

Mean Difference or $\bar{d}$ Before and After Treatment

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

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

where

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

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

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

versus

$$H_1: \delta \neq 0$$

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

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

(For example, see [11].)

Multivariate Extension Paired-Comparison Procedure: p-responses, 2 Treatments and n Experimental Units

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

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

Vector Mean Differences δ Can be Based Upon T^2-Statistic

and the p paired-difference random variables become

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

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

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

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

Specifically,

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

where

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

T^2-Statistic Approximately Distribute as Chi-Square X^2 Random Variable

Specifically,

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

where

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

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

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

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

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

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

Condition δ = 0 Equivalent to "No Average Between the 2 Treatments"

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

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

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

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

Confidence Region for δ Consists of All δ Less Than Standard Deviation

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

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

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

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

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

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

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

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

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

Checking for Mean Difference With Paired Observations

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

Table 6.1 *Effluent Data*

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

Source: Data courtesy of S. Weber.

Checking for Mean Differences With Paired Observations (cont.)

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

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

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

Here

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

and

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

Taking $\alpha = .05$, we find that $[p(n-1)/(n-p)]F_{p,n-p}(.05) = [2(10)/9]F_{2,9}(.05) = 9.47$. Since $T^2 = 13.6 > 9.47$, we reject H_0 and conclude that there is a nonzero mean difference between the measurements of the two laboratories. It appears,

Checking for Mean Differences With Paired Observations (cont.)

mean difference between the measurements of the two laboratories. It appears, from inspection of the data, that the commercial lab tends to produce lower BOD measurements and higher SS measurements than the State Lab of Hygiene. The 95% simultaneous confidence intervals for the mean differences δ_1 and δ_2 can be computed using (6-10). These intervals are

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

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

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

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

Confidence Interval Contains Zero But Ho: δ =0; 95% Simultaneous Confidence Intervals Applies to Entire Set of Intervals Constructed For All Possible Linear Combinations

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

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

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

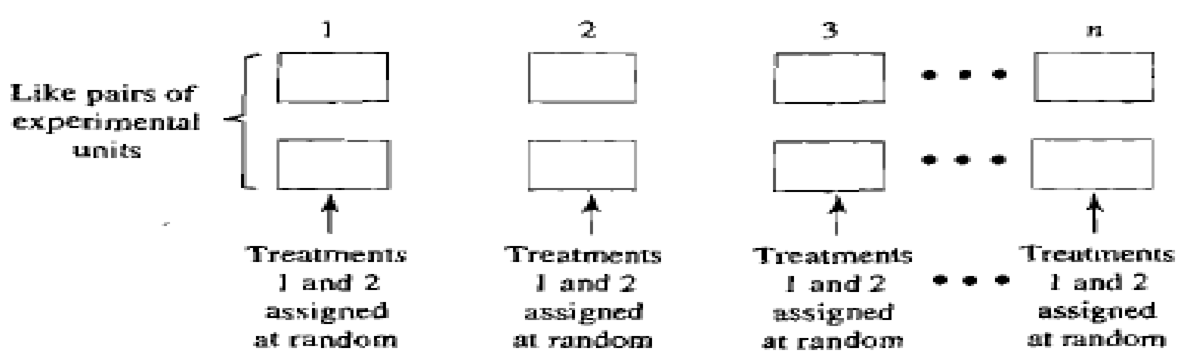
The numerical results of this example illustrate an unusual circumstance that can occur when making inferences. ■

Randomization and Experimental Design for Paired Comparisons

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

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

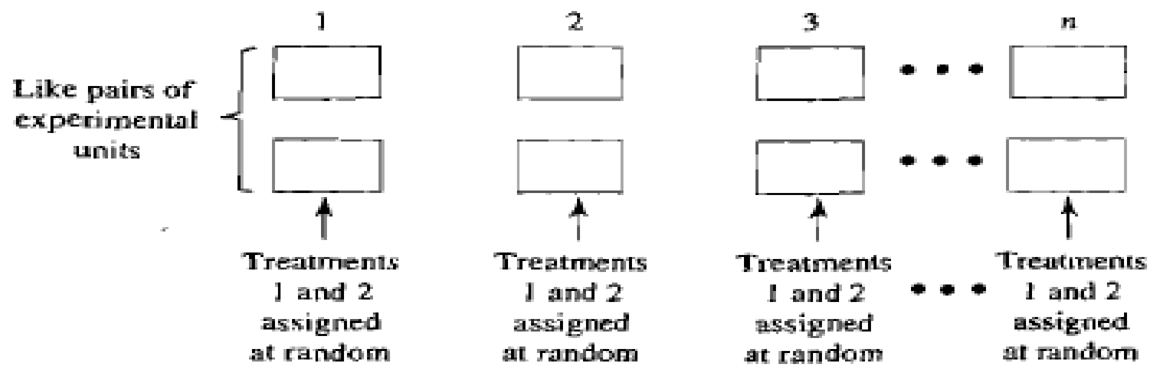
Experimental Design for Paired Comparisons



D-bar and S_d and T^2 Calculated from Full-Sample Quantities x-Bar and S

ceives treatment 1 (heads) or treatment 2 (tails). The remaining treatment is then assigned to the other unit. A separate independent randomization is conducted for each pair. One can conceive of the process as follows:

Experimental Design for Paired Comparisons



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

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

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

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

S_{12} and S_{21} Are the Matrices of Sample Covariance Computed from Observations On Pairs of Treatment 1 and Treatment 2 Variables

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

Defining the matrix

$$\underset{(p\times 2p)}{\mathbf{C}} = \begin{array}{c} \left[\begin{array}{cccc:cccc} 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & -1 \end{array}\right] \\ \uparrow \\ (p+1)\text{st column} \end{array} \tag{6-13}$$

we can verify (see Exercise 6.9) that

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

Thus,

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

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

Each Row c'_i Matrix C is a Contrast Vector, Because Its Elements Sum to Zero

we can verify (see Exercise 6.9) that

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

Thus,

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

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

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

Repeated Measures Design for Comparing Treatments

A Repeated Measures Design for Comparing Treatments

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

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

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

Ho: No Differences in Treatments,i.e. Treatments are Equal or H_o=H_a

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

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

or

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

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

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

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

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

Test for Equality of Treatments in Repeated Measures Design

Test for Equality of Treatments in a Repeated Measures Design

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

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

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

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

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

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

Confidence Region for C_u

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

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

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

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

Testing for Equal Treatments in Repeated Measures Design

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

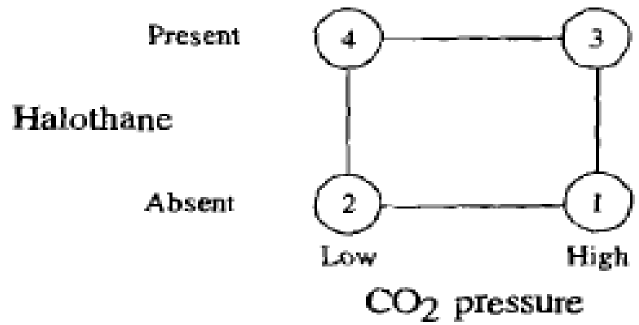


Table 6.2 contains the four measurements for each of the 19 dogs, where

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

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

Three Mean Treatment Contrasts

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

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

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

$$(\mu_1 + \mu_3) - (\mu_2 + \mu_4) = \begin{pmatrix} \text{CO}_2 \text{ contrast representing the difference} \\ \text{between high and low CO}_2 \text{ pressure} \end{pmatrix}$$

$$(\mu_1 + \mu_4) - (\mu_2 + \mu_3) = \begin{pmatrix} \text{Contrast representing the influence} \\ \text{of halothane on CO}_2 \text{ pressure differences} \\ \text{(H–CO}_2 \text{ pressure ``interaction'')} \end{pmatrix}$$

Dogs as Unit of Analysis n = 19 Compared Across 4 Treatments

Table 6.2 Sleeping-Dog Data

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

Source: Data courtesy of Dr. J. Atlee.

Contrast Matrix C, Where Means Are Equal to Each Other

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

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

The data (see Table 6.2) give

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

It can be verified that

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

and

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

Three Sets of Confidence Intervals

With $\alpha = .05$,

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

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

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

is estimated by the interval

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

$$= 209.31 \pm 73.70$$

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

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

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

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

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

Confidence Intervals: Check for Signficance at α = .05

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

$$= 209.31 \pm 73.70$$

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

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

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

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

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

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

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

Comparing Mean Vectors from Two Populations

6.3 Comparing Mean Vectors from Two Populations

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

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

Comparing Mean Vectors from Two Populations: $u_1=u_2$

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

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

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

We want to make inferences about

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

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

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

Assumptions re: Data Structure

Assumptions Concerning the Structure of the Data

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

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

More Assumptions When n_1 and n_2 Small Sample Size

Further Assumptions When n_1 and n_2 Are Small

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

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

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

We set

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

$$= \frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2 \qquad (6\text{-}21)$$

Pooled Variance: When s Approximately Equal

We set

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

$$= \frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2 \qquad (6\text{-}21)$$

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

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

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

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

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

Likelihood Ration Test of Ho: μ_1-μ_2 = δ_0 Based on the Square of the Statistical Distance T^2

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

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

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

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

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

The likelihood ratio test of

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

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

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

Likelihood Ration Test of Ho: μ_1-μ_2 = δ_0 Based on the Square of the Statistical Distance T^2

Reject H_0 if

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

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

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

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

is distributed as

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

Likelihood Ration Test of Ho: μ_1-μ_2 = δ_0 Based on the Square of the Statistical Distance T^2

is distributed as

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

Consequently,

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

where

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

Constructing Confidence Region for the Difference of Two Mean Vectors

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

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

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

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

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

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

Also,

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

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

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

Constructing Confidence Region for the Difference of Two Mean Vectors

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

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

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

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

are

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

By Result 6.2,

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

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

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

95% Confidence Ellipse Not Contained in the Ellipse, Thus Two Methods Producing Bars of Soap Yield Different Results

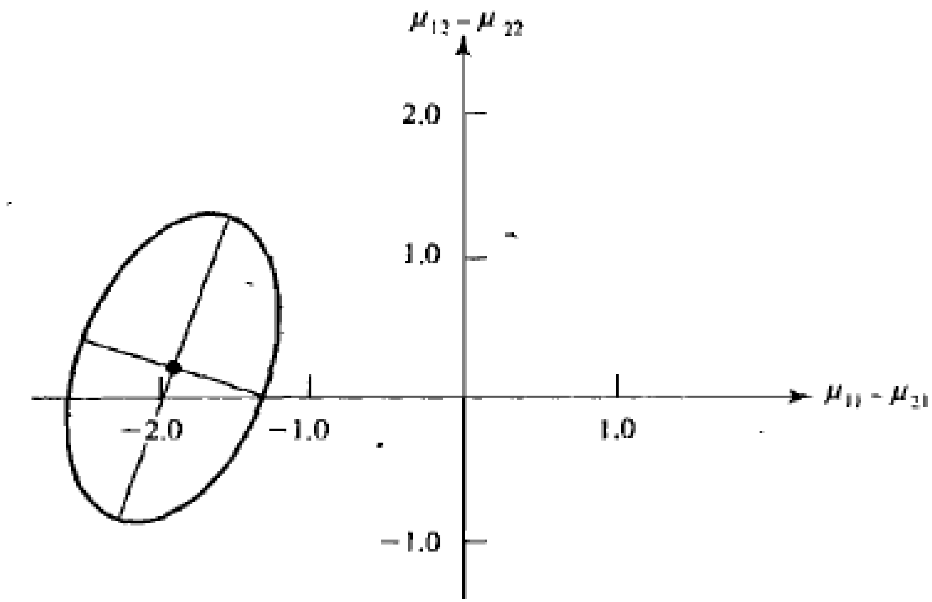


Figure 6.1 95% confidence ellipse for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.

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

Simultaneous Confidence Intervals

Simultaneous Confidence Intervals

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

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

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

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

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

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

Calculating Simultaneous Confidence Intervals for the Differences in Mean Components

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

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

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

Pooled Sample Covariance Matrix

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

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

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

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

and

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

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

Confidence Interval Interpretation: Difference Between Electrical Consumption Those With AC and Those Without

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

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

or

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

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

or

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

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

95% Confidence Ellipse for the 2 Electrical Use Groups

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

Since

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

and

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

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

Confidence Ellipse Does Not Cover [0,0] Then Reject T^2 Statistic Will Reject Ho: μ_1-μ_2 = 0

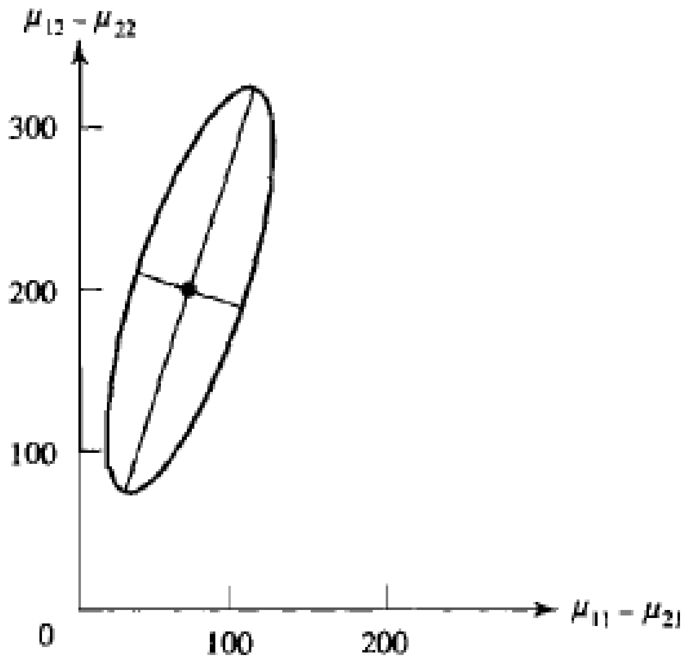


Figure 6.2 95% confidence ellipse for $\boldsymbol{\mu}_1' - \boldsymbol{\mu}_2' = (\mu_{11} - \mu_{21}, \mu_{12} - \mu_{22})$.

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

Bartlett's Test For Equality of $\Sigma_1 \neq \Sigma_2$

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

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

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

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

Bartlett's Test For Equality of $\Sigma_1 \neq \Sigma_2$

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

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

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

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

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

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

and

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

Bartlett's Test For Equality of $\Sigma_1 \neq \Sigma_2$

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

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

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

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

Effect of Unequal Variances is Least When 2 Sample Sizes (Approximately) Equal

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

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

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

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

Confidence Intervals for Large Samples

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

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

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

The 95% simultaneous confidence intervals for the linear combinations

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

and

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

are (see Result 6.4)

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

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

Confidence Intervals for Large Samples, Calculating Hotelling's T^2 Using X^2 Critical Value

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

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

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

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

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

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

For $\alpha = .05$, the critical value is $\chi^2_2(.05) = 5.99$ and, since $T^2 = 15.66 > \chi^2_2(.05) = 5.99$, we reject H_0.

Most Critical Linear Combination Shows Difference in On-Peak Electrical Consumption Between Those With AC and Those Without

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

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

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

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

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

Approximation to Distribution of Hotelling's T^2 For Normal Populations When Sample Sizes Are Not Large

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

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

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

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

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

Degrees of Freedom,v,Estimated from Sample Covariance Matrixes

$$T^2 = \frac{\nu p}{\nu - p + 1} F_{p,\nu-p+1} \tag{6-28}$$

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

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

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

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

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

where the degrees of freedom ν are given by (6-29). This procedure is consistent with the large samples procedure in Result 6.4 except that the critical value $\chi_p^2(\alpha)$ is

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

Approximate Hotelling's T^2 Distribution When of $\Sigma_1 \neq \Sigma_2$

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

We first calculate

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

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

and using a result from Example 6.5,

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

Approximate Hotelling's T^2 Distribution When of $\Sigma_1 \neq \Sigma_2$

Consequently,

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

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

and

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

Further,

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

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

and

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

Approximate Hotelling's T^2 Distribution When of $\Sigma_1 \neq \Sigma_2$

Then

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

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

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

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

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

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

Estimated Degrees of Freedom = 77.6; Calculate F – Ratio, Compare to Critical T^2

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

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

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

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

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

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

Structure of an ANOVA Table

. The summary statistics appear in the last two columns.

TABLE 1 Data Structure for the Completely Randomized Design with *k* Treatments

	Observations	Mean	Sum of Squares
Treatment 1	$y_{11}, y_{12}, \ldots, y_{1n_1}$	$\bar{y}_1$	$\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2$
Treatment 2	$y_{21}, y_{22}, \ldots, y_{2n_2}$	$\bar{y}_2$	$\sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2$
	$\vdots$	$\vdots$	$\vdots$
Treatment k	$y_{k1}, y_{k2}, \ldots, y_{kn_k}$	$\bar{y}_k$	$\sum_{j=1}^{n_k} (y_{kj} - \bar{y}_k)^2$

Grand mean $\bar{y} = \dfrac{\text{Sum of all observations}}{n_1 + n_2 + \cdots + n_k} = \dfrac{n_1\bar{y}_1 + \cdots + n_k\bar{y}_k}{n_1 + \cdots + n_k}$

Before proceeding with the general case of k treatments, it would be instructive to explain the reasoning behind the analysis of variance and the associated calculations in terms of a numerical example.

TABLE 1 Data Structure for the Completely Randomized Design with k Treatments

	Observations	Mean	Sum of Squares
Treatment 1	$y_{11}, y_{12}, \ldots, y_{1n_1}$	$\bar{y}_1$	$\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2$
Treatment 2	$y_{21}, y_{22}, \ldots, y_{2n_2}$	$\bar{y}_2$	$\sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2$
	$\vdots$	$\vdots$	$\vdots$
Treatment k	$y_{k1}, y_{k2}, \ldots, y_{kn_k}$	$\bar{y}_k$	$\sum_{j=1}^{n_k} (y_{kj} - \bar{y}_k)^2$

$$\text{Grand mean } \bar{y} = \frac{\text{Sum of all observations}}{n_1 + n_2 + \cdots + n_k} = \frac{n_1\bar{y}_1 + \cdots + n_k\bar{y}_k}{n_1 + \cdots + n_k}$$

Table 1 (p. 561

Data Structure for the Completely Randomized Design with k Treatments

ANOVA Table: Comparison of Sound Distortion Among 4 Brands Earbuds

Example 1

The Structure of Data from an Experiment for Comparing Four Means

Four models of high-end earbuds are compared for sound reproduction. There are 5 sets of earbud A, 4 of earbud B, 7 of earbud C, and 6 of earbud D available for testing. The quality of sound reproduction can be determined objectively by measuring audio signals received by a robot head wearing earbuds and then comparing them with the known signal wave that was sent. Quantitatively, the measure of sound distortion called total harmonic distortion is an overall measure of the discrepancy, in percent. Because the values are substantially below 1% for high quality earbuds, we give the values for distortion in hundredths of a percent so 10 is .1% and so on. Suppose the test results for sound distortion produce the data in Table 2.

TABLE 2 Sound Distortion Obtained with Four Brands of Earbuds

Earbud	Observations	Mean	Sum of Squares
A	10 15 8 12 15	$\bar{y}_1 = 12$	$\sum_{j=1}^{5} (y_{1j} - \bar{y}_1)^2 = 38$
B	14 18 21 15	$\bar{y}_2 = 17$	$\sum_{j=1}^{4} (y_{2j} - \bar{y}_2)^2 = 30$
C	17 16 14 15 17 15 18	$\bar{y}_3 = 16$	$\sum_{j=1}^{7} (y_{3j} - \bar{y}_3)^2 = 12$
D	12 15 17 15 16 15	$\bar{y}_4 = 15$	$\sum_{j=1}^{6} (y_{4j} - \bar{y}_4)^2 = 14$

Grand mean $\bar{y} = 15$

TABLE 2 Sound Distortion Obtained with Four Brands of Earbuds

Earbud	Observations	Mean	Sum of Squares
A	10 15 8 12 15	$\bar{y}_1 = 12$	$\sum_{j=1}^{5} (y_{1j} - \bar{y}_1)^2 = 38$
B	14 18 21 15	$\bar{y}_2 = 17$	$\sum_{j=1}^{4} (y_{2j} - \bar{y}_2)^2 = 30$
C	17 16 14 15 17 15 18	$\bar{y}_3 = 16$	$\sum_{j=1}^{7} (y_{3j} - \bar{y}_3)^2 = 12$
D	12 15 17 15 16 15	$\bar{y}_4 = 15$	$\sum_{j=1}^{6} (y_{4j} - \bar{y}_4)^2 = 14$
	Grand mean $\bar{y} = 15$		

Table 2 (p. 561)

Sound Distortion Obtained with Four Brands of Earbuds

ANOVA Qx: Does Significant Difference Exist Among the 4 Brands?

TABLE 2 Sound Distortion Obtained with Four Brands of Earbuds

Earbud	Observations	Mean	Sum of Squares
A	10 15 8 12 15	$\bar{y}_1 = 12$	$\sum_{j=1}^{5} (y_{1j} - \bar{y}_1)^2 = 38$
B	14 18 21 15	$\bar{y}_2 = 17$	$\sum_{j=1}^{4} (y_{2j} - \bar{y}_2)^2 = 30$
C	17 16 14 15 17 15 18	$\bar{y}_3 = 16$	$\sum_{j=1}^{7} (y_{3j} - \bar{y}_3)^2 = 12$
D	12 15 17 15 16 15	$\bar{y}_4 = 15$	$\sum_{j=1}^{6} (y_{4j} - \bar{y}_4)^2 = 14$
		Grand mean $\overline{y} = 15$	

Two questions immediately come to mind. Does any significant difference exist among the mean distortions obtained using the four brands of earbuds? Can we establish confidence intervals for the mean differences between brands?

An analysis of the results essentially consists of decomposing the observations into contributions from different sources. We reason that the deviation of an individual observation from the grand mean, $y_{ij} - \overline{y}$, is partly due to differences among the mean qualities of the brands and partly due to random variation in measurements within the

ANOVA Equation: Decomposition of Variation of the Total Plus the Group

An analysis of the results essentially consists of decomposing the observations into contributions from different sources. We reason that the deviation of an individual observation from the grand mean, $y_{ij} - \overline{y}$, is partly due to differences among the mean qualities of the brands and partly due to random variation in measurements within the same group. This suggests the following decomposition.

$$\begin{array}{ccccccc} \text{Observation} & = & \begin{pmatrix}\text{Grand}\\ \text{mean}\end{pmatrix} & + & \begin{pmatrix}\text{Deviation due}\\ \text{to treatment}\end{pmatrix} & + & (\text{Residual}) \\ y_{ij} & = & \overline{y} & + & (\bar{y}_i - \overline{y}) & + & (y_{ij} - \bar{y}_i) \end{array}$$

For the data given in Table 2, the decomposition of all the observations can be presented in the form of the following arrays:

Observations

$$y_{ij}$$

$$\begin{bmatrix} 10 & 15 & 8 & 12 & 15 & & \\ 14 & 18 & 21 & 15 & & & \\ 17 & 16 & 14 & 15 & 17 & 15 & 18 \\ 12 & 15 & 17 & 15 & 16 & 15 & \end{bmatrix}$$

ANOVA Equation: Decomposition of Variation of the Total Plus the Group

Observations

$$y_{ij}$$

$$\begin{bmatrix} 10 & 15 & 8 & 12 & 15 & & \\ 14 & 18 & 21 & 15 & & & \\ 17 & 16 & 14 & 15 & 17 & 15 & 18 \\ 12 & 15 & 17 & 15 & 16 & 15 & \end{bmatrix}$$

Grand mean $\overline{y}$ + **Treatment effects** $(\bar{y}_i - \overline{y})$

$$= \begin{bmatrix} 15 & 15 & 15 & 15 & 15 & & \\ 15 & 15 & 15 & 15 & & & \\ 15 & 15 & 15 & 15 & 15 & 15 & 15 \\ 15 & 15 & 15 & 15 & 15 & 15 & \end{bmatrix} + \begin{bmatrix} -3 & -3 & -3 & -3 & -3 & & \\ 2 & 2 & 2 & 2 & & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \end{bmatrix}$$

Residuals

$$(y_{ij} - \bar{y}_i)$$

$$+ \begin{bmatrix} -2 & 3 & -4 & 0 & 3 & & \\ -3 & 1 & 4 & -2 & & & \\ 1 & 0 & -2 & -1 & 1 & -1 & 2 \\ -3 & 0 & 2 & 0 & 1 & 0 & \end{bmatrix}$$

For instance, the upper left-hand entries of the arrays show that

$$\begin{aligned} 10 &= 15 + (-3) + (-2) \\ y_{11} &= \overline{y} + (\bar{y}_1 - \overline{y}) + \left(y_{11} - \bar{y}_1\right) \end{aligned}$$

ANOVA Equation: Calculation of "Treatment Effects" or "Deviation Due to Treatment" Also Called >Treatment Sum of Squares<

If there is really no difference in the mean distortions obtained using the four brands of earbuds, we can expect the entries of the second array on the right-hand side of the equation, whose terms are $\bar{y}_i - \bar{y}$, to be close to zero. As an overall measure of the amount of variation due to differences in the treatment means, we calculate the sum of squares of all the entries in this array, or

$$\underbrace{(-3)^2 + \cdots + (-3)^2}_{n_1 = 5} + \underbrace{2^2 + \cdots + 2^2}_{n_2 = 4} + \underbrace{1^2 + \cdots + 1^2}_{n_3 = 7} + \underbrace{0^2 + \cdots + 0^2}_{n_4 = 6}$$

$$= 5(-3)^2 + 4(2)^2 + 7(1)^2 + 6(0)^2$$

$$= 68$$

Thus, the sum of squares due to differences in the treatment means, also called the **treatment sum of squares,** is given by

$$\textbf{Treatment sum of squares} = \sum_{i=1}^{4} n_i (\bar{y}_i - \bar{y})^2 = 68$$

ANOVA Equation: Also Known as "Residuals", or Variation Due to Random Error Called >Error Sum of Squares<

The last array consists of the entries $y_{ij} - \bar{y}_i$ that are the deviations of individual observations from the corresponding treatment mean. These deviations reflect inherent variabilities in the material, fabrication, and the measuring device and are called the **residuals**. The overall variation due to random errors is measured by the sum of squares of all these residuals

$$(-2)^2 + 3^2 + (-4)^2 + \cdots + 1^2 + 0^2 = 94$$

Thus, we obtain

$$\textbf{Error sum of squares} = \sum_{i=1}^{4} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = 94$$

The double summation indicates that the elements are summed within each row and then over different rows. Alternatively, referring to the last column in Table 2, we obtain

$$\begin{aligned} \text{Error sum of squares} &= \sum_{j=1}^{5} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{4} (y_{2j} - \bar{y}_2)^2 + \sum_{j=1}^{7} (y_{3j} - \bar{y}_3)^2 + \sum_{j=1}^{6} (y_{4j} - \bar{y}_4)^2 \\ &= 38 + 30 + 12 + 14 = 94 \end{aligned}$$

ANOVA Equation: “Deviations” of Each Observation from Grand Mean Called >Total Sum of Squares<

Finally, the deviations of individual observations from the grand mean $y_{ij} - \overline{y}$ are given by the array

$$\textbf{Deviations} = \begin{bmatrix} -5 & 0 & -7 & -3 & 0 & & \\ -1 & 3 & 6 & 0 & & & \\ 2 & 1 & -1 & 0 & 2 & 0 & 3 \\ -3 & 0 & 2 & 0 & 1 & 0 & \end{bmatrix}$$

The total variation present in the data is measured by the sum of squares of all these deviations.

$$\begin{aligned} \textbf{Total sum of squares} &= \sum_{i=1}^{4}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2 \\ &= (-5)^2 + 0^2 + (-7)^2 + \cdots + 0^2 \\ &= 162 \end{aligned}$$

Note that the total sum of squares is the sum of the treatment sum of squares and the error sum of squares, or,

$$162 = 68 + 94$$

ANOVA: Decomposition of Variance Degrees of Freedom Based on Sums of Squares Because ȳ Weighted Average of the Treatment Means

It is time to turn our attention to another property of this decomposition, the degrees of freedom associated with the sums of squares. In general terms:

$$\begin{pmatrix}\text{Degrees of} \\ \text{freedom} \\ \text{associated with a} \\ \text{sum of squares}\end{pmatrix} = \begin{pmatrix}\text{Number of} \\ \text{elements} \\ \text{whose squares} \\ \text{are summed}\end{pmatrix} - \begin{pmatrix}\text{Number of linear} \\ \text{constraints} \\ \text{satisfied by the} \\ \text{elements}\end{pmatrix}$$

In our present example, the treatment sum of squares is the sum of four terms $n_1(\bar{y}_1-\overline{y})^2+n_2(\bar{y}_2-\overline{y})^2+n_3(\bar{y}_3-\overline{y})^2+n_4(\bar{y}_4-\overline{y})^2$, where the elements satisfy the single constraint

$$n_1(\bar{y}_1-\overline{y})+n_2(\bar{y}_2-\overline{y})+n_3(\bar{y}_3-\overline{y})+n_4(\bar{y}_4-\overline{y})=0$$

This equality holds because the grand mean $\overline{y}$ is a weighted average of the treatment means, or

$$\overline{y}=\frac{n_1\bar{y}_1+n_2\bar{y}_2+n_3\bar{y}_3+n_4\bar{y}_4}{n_1+n_2+n_3+n_4}$$

$$\begin{pmatrix}\text{Degrees of} \\ \text{freedom} \\ \text{associated with a} \\ \text{sum of squares}\end{pmatrix} = \begin{pmatrix}\text{Number of} \\ \text{elements} \\ \text{whose squares} \\ \text{are summed}\end{pmatrix} - \begin{pmatrix}\text{Number of linear} \\ \text{constraints} \\ \text{satisfied by the} \\ \text{elements}\end{pmatrix}$$

Box on Page 564

ANOVA Table For Mean Comparison of Sound Distortion in Earbuds

Consequently, the number of degrees of freedom associated with the treatment sum of squares is $4-1=3$. To determine the degrees of freedom for the error sum of squares, we note that the entries $y_{ij}-\bar{y}_i$ in each row of the residual array sum to zero and there are 4 rows. The number of degrees of freedom for the error sum of squares is then $(n_1+n_2+n_3+n_4)-4=22-4=18$. Finally, the number of degrees of freedom for the total sum of squares is $(n_1+n_2+n_3+n_4)-1=22-1=21$, because the 22 entries $(y_{ij}-\overline{y})$ whose squares are summed satisfy the single constraint that their total is zero. Note that the degrees of freedom for the total sum of squares is the sum of the degrees of freedom for treatment and error, or,

$$21=3+18.$$

We summarize the calculations thus far in Table 3.

TABLE 3 ANOVA Table for Distortion Data

Source	Sum of Squares	d.f.
Treatment	68	3
Error	94	18
Total	162	21

Guided by this numerical example, we now present the general formulas for the analysis of variance for a comparison of k treatments using the data structure given in Table 1. Beginning with the basic decomposition

$$(y_{ij}-\overline{y}) = (\bar{y}_i-\overline{y}) + (y_{ij}-\bar{y}_i)$$

ANOVA – General Formulas Based on Mean Comparison Earbud Distortion

We summarize the calculations thus far in Table 3.

TABLE 3 ANOVA Table for Distortion Data

Source	Sum of Squares	d.f.
Treatment	68	3
Error	94	18
Total	162	21

Guided by this numerical example, we now present the general formulas for the analysis of variance for a comparison of k treatments using the data structure given in Table 1. Beginning with the basic decomposition

$$(y_{ij} - \overline{y}) = (\bar{y}_i - \overline{y}) + (y_{ij} - \bar{y}_i)$$

and squaring each side of the equation, we obtain

$$(y_{ij} - \overline{y})^2 = (\bar{y}_i - \overline{y})^2 + (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \overline{y})(y_{ij} - \bar{y}_i)$$

ANOVA Equation: Total Sum of Squares (SS) Equal to Treatment SS Plus Error SS

Guided by this numerical example, we now present the general formulas for the analysis of variance for a comparison of k treatments using the data structure given in Table 1. Beginning with the basic decomposition

$$(y_{ij} - \overline{y}) = (\bar{y}_i - \overline{y}) + (y_{ij} - \bar{y}_i)$$

and squaring each side of the equation, we obtain

$$(y_{ij} - \overline{y})^2 = (\bar{y}_i - \overline{y})^2 + (y_{ij} - \bar{y}_i)^2 + 2\,(\bar{y}_i - \overline{y})\,(y_{ij} - \bar{y}_i)$$

When summed over $j = 1, \ldots, n_i$ the last term on the right-hand side of this equation reduces to zero due to the relation $\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) = 0$. Therefore, summing each side of the preceding relation over $j = 1, \ldots, n_i$ and $i = 1, \ldots, k$ provides the decomposition

$$\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i} (y_{ij} - \overline{y})^2}_{\text{Total SS}} = \underbrace{\sum_{i=1}^{k} n_i (\bar{y}_i - \overline{y})^2}_{\text{Treatment SS}} + \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}_{\text{Residual SS or error SS}}$$

Total SS	Treatment SS	Residual SS or error SS
$\text{d.f.} = \sum_{i=1}^{k} n_i - 1$	$\text{d.f.} = k - 1$	$\text{d.f.} = \sum_{i=1}^{k} n_i - k$

TABLE 3 ANOVA Table for Distortion Data

Source	Sum of Squares	d.f.
Treatment	68	3
Error	94	18
Total	162	21

Table 3 (p. 564)

ANOVA Table for Distortion Data

ANOVA: Calculation of Total SS Based on Degrees of Freedom for Group and Error

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\overline{y})^2 = \sum_{i=1}^{k} n_i(\bar{y}_i-\overline{y})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$$

↑	↑	↑
Total SS	**Treatment SS**	**Residual SS or error SS**
$\text{d.f.} = \sum_{i=1}^{k} n_i - 1$	$\text{d.f.} = k-1$	$\text{d.f.} = \sum_{i=1}^{k} n_i - k$

We summarize the two key results.

$$\begin{aligned}\textbf{Total SS} &= \textbf{Treatment SS} + \textbf{Error SS}\\ \textbf{d.f.}_{\textbf{Total}} &= \textbf{d.f.}_{\textbf{Treatment}} + \textbf{d.f.}_{\textbf{Error}}\end{aligned}$$

It is customary to present the decomposition of the sum of squares and the degrees of freedom in a tabular form called the **analysis of variance table**, abbreviated as **ANOVA table**. Table 4 contains the additional column for the **mean square** associated with a component, which is defined as

$$\text{Mean square} = \frac{\text{Sum of squares}}{\text{d.f.}}$$

ANOVA: Degrees of Freedom Used Due to F Distribution

$$\begin{array}{lclcl} \textbf{Total SS} & = & \textbf{Treatment SS} & + & \textbf{Error SS} \\ \textbf{d.f.}_{\textbf{Total}} & = & \textbf{d.f.}_{\textbf{Treatment}} & + & \textbf{d.f.}_{\textbf{Error}} \end{array}$$

It is customary to present the decomposition of the sum of squares and the degrees of freedom in a tabular form called the **analysis of variance table**, abbreviated as **ANOVA table**. Table 4 contains the additional column for the **mean square** associated with a component, which is defined as

$$\text{Mean square} = \frac{\text{Sum of squares}}{\text{d.f.}}$$

In the ANOVA Table for comparing k treatments, we set $n = \sum_{i=1}^{k} n_i$.

TABLE 4 ANOVA Table for Comparing k Treatments

Source	Sum of Squares	d.f.	Mean Square
Treatment	$SS_T = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2$	$k - 1$	$MS_T = \frac{SS_T}{k-1}$
Error	$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$n - k$	$MSE = \frac{SSE}{n-k}$
Total	$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	$n - 1$	

TABLE 4 ANOVA Table for Comparing k Treatments

Source	Sum of Squares	d.f.	Mean Square
Treatment	$SS_T = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2$	$k - 1$	$MS_T = \dfrac{SS_T}{k-1}$
Error	$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$	$\sum_{i=1}^{k} n_i - k$	$MSE = \dfrac{SSE}{\sum_{i=1}^{k} n_i - k}$
Total	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2$	$\sum_{i=1}^{k} n_i - 1$	

Table 4 (p. 565)

ANOVA Table for Comparing k Treatments

ANOVA: Calculation Without Computer

1.1 GUIDE TO HAND CALCULATION

When performing an ANOVA on a calculator, it is convenient to express the sums of squares in an alternative form. These employ the treatment totals

$$T_i = \sum_{j=1}^{n_i} y_{ij} = \text{Sum of all responses under treatment } i$$

$$T = \sum_{i=1}^{k} T_i = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = \text{Sum of all observations}$$

ANOVA Calculation Without Computer

.

to calculate the sums of squares:

$$\text{Total SS} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{n} \quad \text{where} \quad n = \sum_{i=1}^{k} n_i$$

$$SS_T = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - \frac{T^2}{n}$$

$$SSE = \text{Total SS} - SS_T$$

Notice that the SSE can be obtained by subtraction.

Example: ANOVA Calculation

Example 2

Calculating Sums of Squares Using the Alternative Formulas

Obtain the Total SS, SS_T, and SSE for the data in Example 1 using the alternative form of calculation.

SOLUTION

$$
\begin{aligned}
T_1 &= 10+15+8+12+15=60 & n_1 &= 5 \\
T_2 &= 14+18+21+15=68 & n_2 &= 4 \\
T_3 &= 17+16+14+15+17+15+18=112 & n_3 &= 7 \\
T_4 &= 12+15+17+15+16+15=90 & n_4 &= 6
\end{aligned}
$$

and

$$
\begin{aligned}
T &= T_1+T_2+T_3+T_4 & n &= n_1+n_2+n_3+n_4 \\
&= 60+68+112+90=330 & &= 5+4+7+6=22
\end{aligned}
$$

Example: ANOVA Calculation

Since

$$\sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}^2 = (10)^2 + (15)^2 + \cdots + (16)^2 + (15)^2 = 5112$$

$$\text{Total SS} = 5112 - \frac{(330)^2}{22} = 162$$

$$SS_T = \frac{(60)^2}{5} + \frac{(68)^2}{4} + \frac{(112)^2}{7} + \frac{(90)^2}{6} - \frac{(330)^2}{22} = 68$$

$$\text{SSE} = \text{Total SS} - SS_T = 162 - 68 = 94$$

ANOVA – Common Variance with Mutually Independent Samples

2. Population Model and Inferences for a One-Way Analysis of Variance

To implement a formal statistical test for no difference among treatment effects, we need to have a population model for the experiment. To this end, we assume that the response measurements with the ith treatment constitute a random sample from a normal population with a mean of μ_i and a **common variance** of σ^2. The samples are assumed to be mutually independent.

Population Model for Comparing *k* Treatments

$$Y_{ij} = \mu_i + e_{ij} \quad j = 1, \ldots, n_i \ \text{ and } \ i = 1, \ldots, k$$

where $\mu_i = i$th treatment mean. The errors e_{ij} are all independently distributed as $N(0, \sigma)$.

Population Model for Comparing k Treatments

$$Y_{ij} = \mu_i + e_{ij} \qquad j = 1, \ldots, n_i \qquad \text{and} \qquad i = 1, \ldots, k$$

where $\mu_i = i$th treatment mean. The errors e_{ij} are all independently distributed as $N(0, \sigma)$.

Box on Page 568

Population Model for Comparing k Treatments

ANOVA Principle: For each mean or μ of the group = Overall Effect μ + the Effect Due to The ith Treatment a_i

Before presenting the test for equality of means based on this model, we give a second parametrization that previews the formulation of statistical models for more complicated designs. For each i, the mean μ_i is considered to be the sum of an overall effect μ, common to all treatments, and an effect due only to the ith treatment.

$$\mu_i = \mu + \alpha_i \quad \text{subject to the constraint} \sum_{i=1}^{k} n_i \alpha_i = 0$$

The right-hand side, $\mu + \alpha_i$, is estimated by $\overline{y} + (\bar{y}_i - \overline{y})$ in the decomposition and the estimated treatment effects satisfy the same constraint as the α_i's.

F Test – Detecting Difference Between Means of Groups

2.1 *F* Distribution

The F test will determine if significant differences exist between the k sample means. The null hypothesis that no difference exists among the k population means can now be phrased as follows:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

The alternative hypothesis is that not all the μ_i's are equal. Seeking a criterion to test the null hypothesis, we observe that when the population means are all equal, $\bar{y}_i - \overline{y}$ is expected to be small, and consequently, the treatment mean square $\sum n_i (\bar{y}_i - \overline{y})^2 / (k-1)$ is expected to be small. On the other hand, it is likely to be large when the means differ markedly. The error mean square, which provides an estimate of σ^2, can be used as a yardstick for determining how large a treatment mean square should be before it indicates significant differences. Its value is not influenced by differences in the treatment means.

F Test Ratio and F Distribution

Its value is not influenced by differences in the treatment means.

Statistical distribution theory tells us that under H_0 the ratio

$$F = \frac{\text{Treatment mean square}}{\text{Error mean square}} = \frac{\text{Treatment SS} / (k-1)}{\text{Error SS} / (n-k)}$$

has an ***F* distribution** with d.f. $= (k-1, n-k)$, where $n = \sum n_i$.

Notice that an F distribution is specified in terms of its numerator degrees of freedom $v_1 = k-1$ and denominator degrees of freedom $v_2 = n-k$. We denote

$$F_\alpha(v_1, v_2) = \text{Upper } \alpha \text{ point of the } F \text{ distribution with } (v_1, v_2) \text{ d.f.}$$

which is also called the upper 100α-th percentage point.

F Test for Equality of Means

Reject $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ if

$$F = \frac{\text{Treatment SS}/(k - 1)}{\text{Error SS}/(n - k)} \geq F_\alpha(k - 1, n - k)$$

where $n = \sum_{i=1}^{k} n_i$ and $F_\alpha(k - 1, n - k)$ is the upper α point of the F distribution with d.f. $= (k - 1, n - k)$.

Box on Page 570

F Test for Equality of Means

Reading the F DistributionTable

The upper $\alpha = .05$ and $\alpha = .10$ points are given in Appendix B, Table 6, for several pairs of d.f. With $\nu_1 = 7$ and $\nu_2 = 15$, for $\alpha = .05$, we read from column $\nu_1 = 7$ and row $\nu_2 = 15$ to obtain $F_{.05}(7, 15) = 2.71$ (see Table 5).

TABLE 5 Percentage Points of $F(\nu_1, \nu_2)$ Distributions $\alpha = .05$

ν_2 \ ν_1	. . .	7	. . .
.		.	
.		.	
.		.	
15	. . .	2.71	
.			
.			
.			

We summarize the F test introduced above.

Table 5 (p. 570)

Percentage Points of (F(v_1, v_2) Distributions α = 5

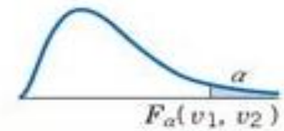
TABLE 5 Percentage Points of $F(v_1, v_2)$ Distributions $\alpha = .05$

v_2 \ v_1	⋯	7	⋯
⋮		⋮	
15	⋯	2.71	
⋮			

TABLE 7 Percentage Points of $F(v_1, v_2)$ Distributions

F Test Table

$\alpha = .05$



v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0	249.3	250.1	251.1	252.2
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.57
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.43
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.74
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.30
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.01
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.79
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.62
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.49
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.38
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.30
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.22
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.16
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.11
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.06
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.02
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.95
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.02	1.98	1.94	1.89
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.86
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.94	1.90	1.85	1.80
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.92	1.88	1.84	1.79
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.89	1.85	1.81	1.75
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.74
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.64
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.69	1.65	1.59	1.53
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.60	1.55	1.50	1.43
∞	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.51	1.46	1.39	1.32

F Test for Equality of Means Formula

We summarize the F test introduced above.

F Test for Equality of Means

Reject $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ if

$$F = \frac{\text{Treatment SS} / (k-1)}{\text{Error SS} / (n-k)} \geq F_\alpha (k-1, n-k)$$

where $n = \sum_{i=1}^{k} n_i$ and $F_\alpha (k-1, n-k)$ is the upper α point of the F distribution with d.f. $= (k-1, n-k)$.

The computed value of the F-ratio is usually presented in the last column of the ANOVA table.

F Test Example: Sound Distortion in Four Brands of Earbuds

Example 3

The F Test for Testing the Null Hypothesis of No Difference in Sound Distortion Means

Construct the ANOVA table for the data given in Example 1 concerning a comparison of four brands of earbuds. Test the null hypothesis that the means are equal. Use $\alpha = .05$.

SOLUTION

Using our earlier calculations for the component sums of squares, we construct the ANOVA table that appears in Table 6.

A test of the null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ is performed by comparing the observed F value 4.34 with the tabulated value of F with d.f. $= (3, 18)$. At a .05 level of significance, the tabulated value is found to be 3.16. Because this is exceeded by the observed value, we conclude that there is a significant difference among the four mean sound distortions. A computer calculation gives P-value $= .018$, as illustrated in Figure 1, which further strengthens the conclusion of unequal mean distortions for the earbuds.

Example 1 Data: Sound Distortion in Four Brands of Earbuds

TABLE 2 Sound Distortion Obtained with Four Brands of Earbuds

Earbud	Observations	Mean	Sum of Squares
A	10 15 8 12 15	$\bar{y}_1 = 12$	$\sum_{j=1}^{5} (y_{1j} - \bar{y}_1)^2 = 38$
B	14 18 21 15	$\bar{y}_2 = 17$	$\sum_{j=1}^{4} (y_{2j} - \bar{y}_2)^2 = 30$
C	17 16 14 15 17 15 18	$\bar{y}_3 = 16$	$\sum_{j=1}^{7} (y_{3j} - \bar{y}_3)^2 = 12$
D	12 15 17 15 16 15	$\bar{y}_4 = 15$	$\sum_{j=1}^{6} (y_{4j} - \bar{y}_4)^2 = 14$
		Grand mean $\overline{y} = 15$	

ANOVA Table for Sound Distortion in Four Brands of Earbuds

TABLE 6 ANOVA Table for the Data Given in Example 1

Source	Sum of Squares	d.f.	Mean Square	F-ratio
Treatment	68	3	22.67	$\frac{22.67}{5.22} = 4.34$
Error	94	18	5.22	
Total	162	21		

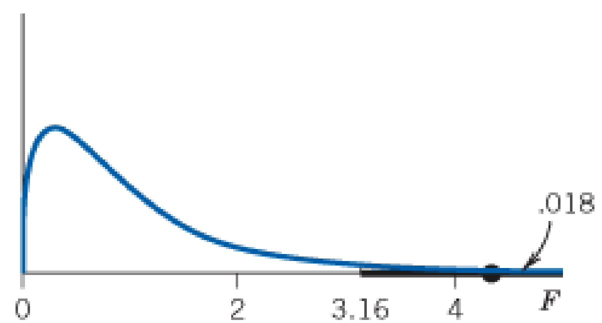


Figure 1 The rejection region and P-value for the F test in Example 3.

TABLE 6 ANOVA Table for the Data Given in Example 1

Source	Sum of Squares	d.f.	Mean Square	*F*-ratio
Treatment	68	3	22.67	$\frac{22.67}{5.22} = 4.34$
Error	94	18	5.22	
Total	162	21		

Table 6 (p. 571)

ANOVA Table for the Data Given in Example 1

ANOVA Computer Output for Sound Distortion of Four Brands Earbuds

Creating the ANOVA table via computer software is the best method and the P-value is included.

Table 7 gives some typical output from a computer program where the term factor is used instead of treatment. The MINITAB commands for obtaining these results are given in Exercise 14.37.

TABLE 7 Computer Output: One-Way Analysis of Variance for Sound Distortion Data

One-way Analysis of Variance

Analysis of Variance

Source	DF	SS	MS	F	P
Factor	3	68.00	22.67	4.34	0.018
Error	18	94.00	5.22		
Total	21	162.00			

TABLE 7 Computer Output: One-Way Analysis of Variance for Distortion Data

```
One-way Analysis of Variance

Analysis of Variance
Source      DF        SS        MS        F        P
Factor       3     68.00     22.67     4.34    0.018
Error       18     94.00      5.22
Total       21    162.00
```

Table 5.7 (p. 571)

Computer Output: One-Way Analysis of Variance for Distortion Data

Matched Pair Comparisons Study Design: Used to Ensure Treatment Efficacy or Effectiveness

5. Matched Pairs Comparisons

In comparing two treatments, it is desirable that the experimental units or subjects be as alike as possible, so that a difference in responses between the two groups can be attributed to differences in treatments. If some identifiable conditions vary over the units in an uncontrolled manner, they could introduce a large variability in the measurements. In turn, this could obscure a real difference in treatment effects. On the other hand, the requirement that all subjects be alike may impose a severe limitation on the number of subjects available for a comparative experiment. To compare two analgesics, for example, it would be impractical to look for a sizable number of patients who are of the same sex, age, and general health condition and who have the same severity of pain. Aside from the question of practicality, we would rarely want to confine a comparison to such a narrow group. A broader scope of inference can be attained by applying the treatments on a variety of patients of both sexes and different age groups and health conditions.

Matching or Blocking Fundamental Element of Matched Pairs Study Design

The concept of **matching** or **blocking** is fundamental to providing a compromise between the two conflicting requirements that the experimental units be alike and also of different kinds. The procedure consists of choosing units in pairs or blocks so that the units in each block are similar and those in different blocks are dissimilar. One of the units in each block is assigned to treatment 1, the other to treatment 2. This process preserves the effectiveness of a comparison within each block and permits a diversity of conditions to exist in different blocks. Of course, the treatments must be allotted to each pair randomly to avoid selection bias. This design is called a **matched pairs design** or **sampling**. For example, in studying how two different environments influence the learning capacities of preschoolers, it is desirable to remove the effect of heredity: Ideally, this is accomplished by working with twins.

Matched Pairs Design

Matched pair	Experimental units	
1	2	1
2	1	2
3	1	2
⋮	⋮	⋮
n	2	1

Units in each pair are alike, whereas units in different pairs may be dissimilar. In each pair, a unit is chosen at random to receive treatment 1, the other unit receives treatment 2.

Matched Pairs Design

Matched pair	Experimental units	
1	2	1
2	1	2
3	1	2
⋮	⋮	⋮
n	2	1

Units in each pair are alike, whereas units in different pairs may be dissimilar. In each pair, a unit is chosen at random to receive treatment 1, the other unit receives treatment 2.

Box on Page 419

Matched Pairs Design

Detecting Differences in Matched Pair Study Design: Use Paired-Difference Formula

The structure of the observations in a paired comparison is given below, where X and Y denote the responses to treatments 1 and 2, respectively. The difference between the responses in each pair is recorded in the last column, and the summary statistics are also presented.

Structure of Data for a Matched Pair Comparison

Pair	Treatment 1	Treatment 2	Difference
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_2	Y_2	$D_2 = X_2 - Y_2$
.	.	.	.
.	.	.	.
.	.	.	.
n	X_n	Y_n	$D_n = X_n = Y_n$

The differences $D_1, D_2, \ldots, D_n$ are a random sample.

Summary statistics:

$$\overline{D} = \frac{1}{n} \sum_{i=1}^{n} D_i \quad s_D^2 = \frac{\sum_{i=1}^{n} (D_i - \overline{D})^2}{n-1}$$

Basic Assumption of Paired-Difference: If Mean Difference Equals Zero Then The Two Treatments are Equivalent

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \quad S_D^2 = \frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1}$$

Although the pairs (X_i, Y_i) are independent of one another, X_i and Y_i within the ith pair will usually be dependent. In fact, if the pairing of experimental units is effective, we would expect X_i and Y_i to be relatively large or small together. Expressed in another way, we would expect (X_i, Y_i) to have a high positive correlation. Because the differences $D_i = X_i - Y_i$, $i = 1, 2, \ldots, n$ are freed from the block effects, it is reasonable to assume that they constitute a random sample from a population with mean μ_D and variance σ_D^2, where μ_D represents the mean difference of the treatment effects. In other words,

$$E(D_i) = \mu_D \quad \text{and} \quad \mathrm{Var}(D_i) = \sigma_D^2 \text{ for } i = 1, \ldots, n$$

If the mean difference μ_D is zero, then the two treatments can be considered equivalent. A positive μ_D signifies that treatment 1 has a higher mean response than treatment 2. Considering $D_1, \ldots, D_n$ to be a single random sample from a population, we can immediately apply the techniques discussed in Chapters 8 and 9 to learn about the population mean μ_D.

Pairing (or Blocking)

Pairing like experimental units according to some identifiable characteristic(s) serves to remove this source of variation from the experiment.

Structure of Data for a Matched Pair Comparison

Pair	Treatment 1	Treatment 2	Difference
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_2	Y_2	$D_2 = X_2 - Y_2$
.	.	.	.
.	.	.	.
.	.	.	.
n	X_n	Y_n	$D_n = X_n - Y_n$

The differences $D_1, D_2, \ldots, D_n$ are a random sample.
Summary statistics:

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \qquad S_D^2 = \frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1}$$

Boxes on Page 420

Pairing (or Blocking; Structure of Data for a Matched Pair Comparison

One-Way MANOVA (Several Multivariate Population Means)

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

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

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

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

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

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

Assumptions of Data Structure for One-Way MANOVA

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

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

2. All populations have a common covariance matrix $\boldsymbol{\Sigma}$.

3. Each population is multivariate normal.

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

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

Summary Univariate ANOVA

A Summary of Univariate ANOVA

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

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

The *reparameterization*

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

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

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

Summary Univariate ANOVA

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

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

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

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

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

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

Sum of Squares Decomposition Univariate ANOVA

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

Population 1: 9, 6, 9

Population 2: 0, 2

Population 3: 3, 1, 2

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

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

Repeating this operation for each observation, we obtain the arrays

$$\underset{\substack{\text{observation} \\ (x_{\ell j})}}{\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & \\ 3 & 1 & 2 \end{pmatrix}} = \underset{\substack{\text{mean} \\ (\bar{x})}}{\begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 & \\ 4 & 4 & 4 \end{pmatrix}} + \underset{\substack{\text{treatment effect} \\ (\bar{x}_\ell - \bar{x})}}{\begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 & \\ -2 & -2 & -2 \end{pmatrix}} + \underset{\substack{\text{residual} \\ (x_{\ell j} - \bar{x}_\ell)}}{\begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & \\ 1 & -1 & 0 \end{pmatrix}}$$

The question of equality of means is answered by assessing whether the contribution of the treatment array is large relative to the residuals. (Our estimates $\hat{\tau}_\ell = \bar{x}_\ell - \bar{x}$ of τ_ℓ always satisfy $\sum_{\ell=1}^{g} n_\ell \hat{\tau}_\ell = 0$. Under H_0, each $\hat{\tau}_\ell$ is an estimate of zero.) If the treatment contribution is large, H_0 should be rejected. The

Sum of Squares Decomposition Univariate ANOVA

estimate of zero.) If the treatment contribution is large, H_0 should be rejected. The size of an array is quantified by stringing the rows of the array out into a vector and calculating its squared length. This quantity is called the *sum of squares* (SS). For the observations, we construct the vector $\mathbf{y}' = [9, 6, 9, 0, 2, 3, 1, 2]$. Its squared length is

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

Similarly,

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

and the residual sum of squares is

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

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

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

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

Sum of Squares Decomposition Univariate ANOVA

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

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

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

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

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

Next, summing both sides over ℓ we get

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

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

or

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

Vector Representations of Arrays In the Decomposition Have Geometric Interpretations that Provide d.f.

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

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

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

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

Mean Vector Lies Anywhere Along the One-Dimensional Equiangular Line; Treatment Vector Lies Anywhere in the Other g-1 Dimensions; Residual Vector Perpendicular to the Mean and Treatment Vectors

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

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

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

ANOVA Table for Comparing Univariate Population Means

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

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

ANOVA Table for Comparing Univariate Population Means

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

F-Test to Reject Hypothesis Treatment Means Equal

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

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

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

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

Example: Univariate ANOVA and F-Test

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

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

Consequently,

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

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

MANOVA Model for Comparing g Population Mean Vectors

MANOVA Model For Comparing g Population Mean Vectors

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

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

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

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

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

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

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

Decomposition of Observation Vectors

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

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

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

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

can be written as

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

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

$$\begin{array}{ccccc} \displaystyle\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})' & = & \displaystyle\sum_{\ell=1}^{g} n_\ell (\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})' & + & \displaystyle\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)' \\ \begin{pmatrix}\text{total (corrected) sum}\\ \text{of squares and cross}\\ \text{products}\end{pmatrix} & & \begin{pmatrix}\text{treatment (\underline{B}etween)}\\ \text{sum of squares and}\\ \text{cross products}\end{pmatrix} & & \begin{pmatrix}\text{residual (\underline{W}ithin) sum}\\ \text{of squares and cross}\\ \text{products}\end{pmatrix} \end{array} \tag{6-40}$$

Total Sum of Squares = Treatment Plus Residual Sum of Squares and Cross-Products

$$\begin{pmatrix}\text{total (corrected) sum}\\ \text{of squares and cross}\\ \text{products}\end{pmatrix} \quad \begin{pmatrix}\text{treatment (\underline{B}etween)}\\ \text{sum of squares and}\\ \text{cross products}\end{pmatrix} \quad \begin{pmatrix}\text{residual (\underline{W}ithin) sum}\\ \text{of squares and cross}\\ \text{products}\end{pmatrix} \tag{6-40}$$

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

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

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

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

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

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

MANOVA for Comparing Population Mean Vectors

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

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

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

MANOVA Table for Comparing Population Mean Vectors

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

MANOVA Table Same as ANOVA

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

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

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

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

Using Wilks Lambda to Test Ratio of Variances: If Ratio Too Small Reject

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

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

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

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

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

Calculated Wilks' Lambda Greater than F Chi-Square Distribution Then Reject

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

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

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

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

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

MANOVA and Wilks' Lambda for Testing Equality of Three Mean Vectors

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

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

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

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

and

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

MANOVA and Wilks' Lambda for Testing Equality of Three Mean Vectors

overall mean, treatment effect, and residual in our discussion of univariate ANOVA. We found that

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

(observation) (mean) (treatment effect) (residual)

and

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

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

$$\begin{pmatrix} 3 & 2 & 7 \\ 4 & 0 & \\ 8 & 9 & 7 \end{pmatrix} = \begin{pmatrix} 5 & 5 & 5 \\ 5 & 5 & \\ 5 & 5 & 5 \end{pmatrix} + \begin{pmatrix} -1 & -1 & -1 \\ -3 & -3 & \\ 3 & 3 & 3 \end{pmatrix} + \begin{pmatrix} -1 & -2 & 3 \\ 2 & -2 & \\ 0 & 1 & -1 \end{pmatrix}$$

(observation) (mean) (treatment effect) (residual)

MANOVA and Wilks' Lambda for Testing Equality of Three Mean Vectors

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

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

and

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

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

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

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

MANOVA Table Format

Thus, the MANOVA table takes the following form:

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

Equation (6-40) is verified by noting that

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

Using (6-42), we get

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

MANOVA and Wilks' Lambda for Testing Equality of Three Mean Vectors

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

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

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

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

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

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

MANOVA of WI Nursing Home Data

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

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

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

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

MANOVA of WI Nursing Home Data

Sample covariance matrices

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

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

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

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

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

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

Also,

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

MANOVA of WI Nursing Home Data

Also,

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

and

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

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

Computer-based calculations give

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

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

MANOVA of WI Nursing Home Data

and

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

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

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

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

with $\chi^2_{p(g-1)}(.01) = \chi^2_8(.01) = 20.09$. Since $132.76 > \chi^2_8(.01) = 20.09$, we reject H_0 at the 1% level. This result is consistent with the result based on the foregoing F-statistic. ■

Simultaneous Confidence Intervals for Treatment Effects

6.5 Simultaneous Confidence Intervals for Treatment Effects

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

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

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

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

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

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

Error Rate Apportioned Over the Numerous Confidence Statements

That is,

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

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

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

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

is the number of simultaneous confidence statements.

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

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

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

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

Simultaneous Intervals for Treatment Differences in WI Nursing Homes

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

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

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

Consequently,

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

Simultaneous Intervals for Treatment Differences in WI Nursing Homes

Consequently,

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

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

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

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

$$\tau_{13} - \tau_{33} \text{ belongs to } \hat{\tau}_{13} - \hat{\tau}_{33} \pm t_{513}(.00208)\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_3}\right)\frac{w_{33}}{n-g}}$$

$$= -.043 \pm 2.87(.00614)$$

$$= -.043 \pm .018, \text{ or } (-.061, -.025)$$

Difference in Cost Exists Between Private and Nonprofit Nursing, But No Difference Between Nonprofit and Government Nursing Homes

$$\tau_{13} - \tau_{33} \text{ belongs to } \hat{\tau}_{13} - \hat{\tau}_{33} \pm t_{513}(.00208)\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_3}\right)\frac{w_{33}}{n-g}}$$

$$= -.043 \pm 2.87(.00614)$$

$$= -.043 \pm .018, \text{ or } (-.061, -.025)$$

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

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

and

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

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

Testing for Equality of Covariance Matrices

6.6 Testing for Equality of Covariance Matrices

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

With g populations, the null hypothesis is

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

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

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

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

Pooled Sample Covariance Matrix

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

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

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

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

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

Box's Test for Equality of Covariance Matrices

Box's Test for Equality of Covariance Matrices

Set

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

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

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

has an approximate χ^2 distribution with

$$\nu = g\frac{1}{2}p(p+1) - \frac{1}{2}p(p+1) = \frac{1}{2}p(p+1)(g-1) \qquad (6\text{-}53)$$

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

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

Testing Equality of Covariance Matrices With WI Nursing Home Data

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

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

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

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

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

Box's M-Test (Using Chi-Square as Critical Value) Sensitive to Non-Normality

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

Experimental Design and Randomization Review

Experimental Design: Independent Samples (Complete Randomization) vs. Matched Pairs Sample: Randomization with Each Match Pair

1. Two Designs: Independent Samples and Matched Pairs Sample

When discussing a comparative study, the common statistical term **treatment** is used to refer to the things that are being compared. The basic unit that is exposed to one treatment or another is called an **experimental unit** or **experimental subject**, and the characteristic that is recorded after the application of a treatment to a subject is called the **response**. For instance, the two treatments in Example 1 are the current and new pesticides, the experimental subjects are the agricultural plots, and the response is the amount of residue in vegetables.
The term **experimental design** refers to the manner in which subjects are chosen and assigned to treatments. For comparing two treatments, the two basic types of design are:

1. **Independent samples (complete randomization)**.
2. **Matched pairs sample (randomization within each matched pair)**.

Selection of Independent Samples vs. Matched Pair Samples

The case of independent samples arises when the subjects are randomly divided into two groups, one group is assigned to treatment 1 and the other to treatment 2. The response measurements for the two treatments are then unrelated because they arise from separate and unrelated groups of subjects. Consequently, each set of response measurements can be considered a sample from a population, and we can speak in terms of a comparison between two population distributions.

With the matched pairs design, the experimental subjects are chosen in pairs so that the members in each pair are alike, whereas those in different pairs may be substantially dissimilar. One member of each pair receives treatment 1 and the other treatment 2. Example 3 illustrates these ideas.

Illustration: Independent Samples vs. Matched Pair Sample Design

Example 3

Independent Samples versus Matched Pairs Design

To compare the effectiveness of two drugs in curing a disease, suppose that 8 patients are included in a clinical study. Here, the time to cure is the response of interest. Figure 1a portrays a design of independent samples where the 8 patients are randomly split into groups of 4, one group is treated with drug 1, and the other with drug 2. The observations for drug 1 will have no relation to those for drug 2 because the selection of patients in the two groups is left completely to chance.

To conduct a matched pairs design, one would first select the patients in pairs. The two patients in each pair should be as alike as possible in regard to their physiological conditions; for instance, they should be of the same gender and age group and have about the same severity of the disease. These preexisting conditions may be different from one pair to another. Having paired the subjects, we randomly select one member from each pair to be treated with drug 1 and the other with drug 2. Figure 1b shows this matched pairs design.

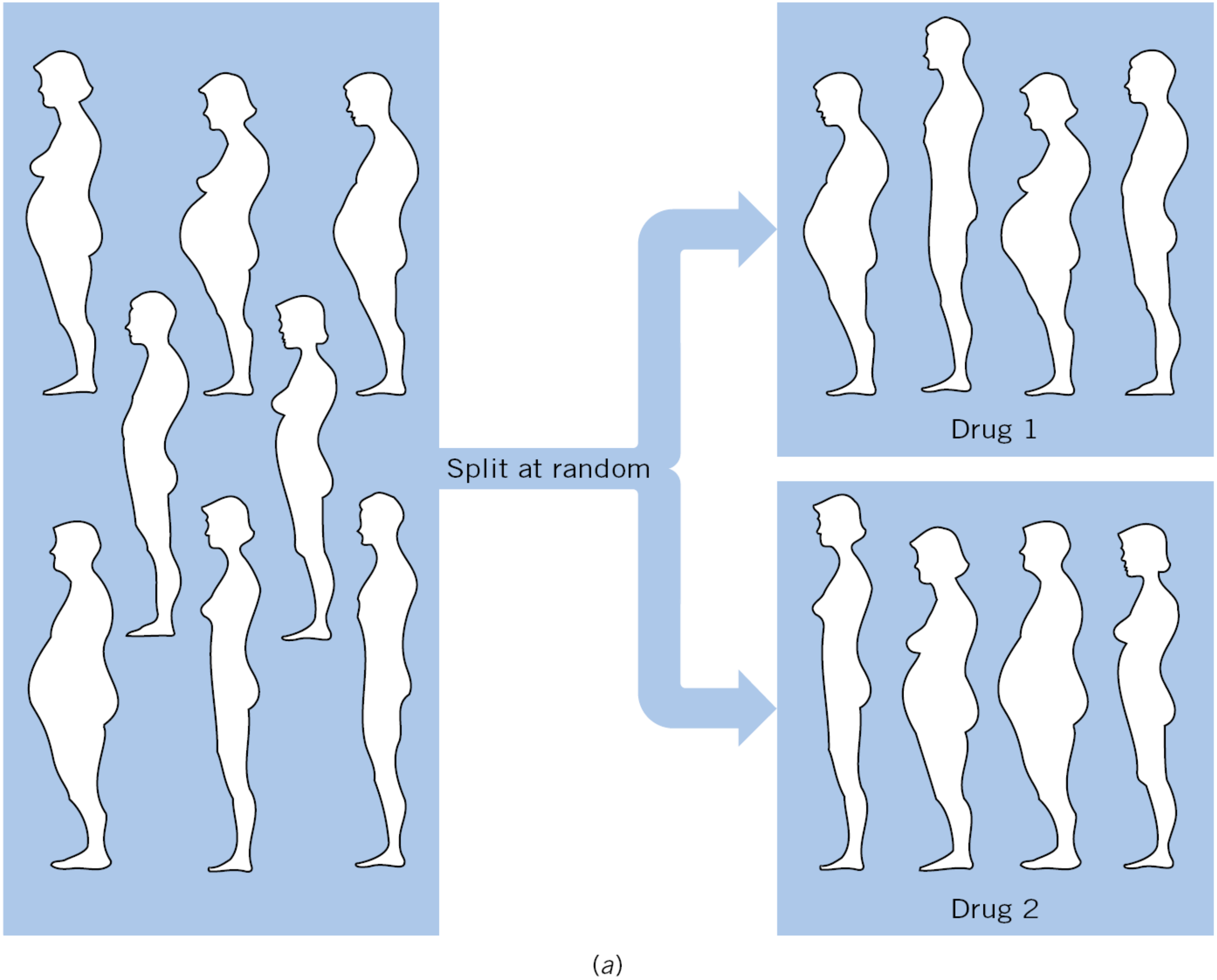


Figure 1a (p. 393)

Independent samples, each of size 4.

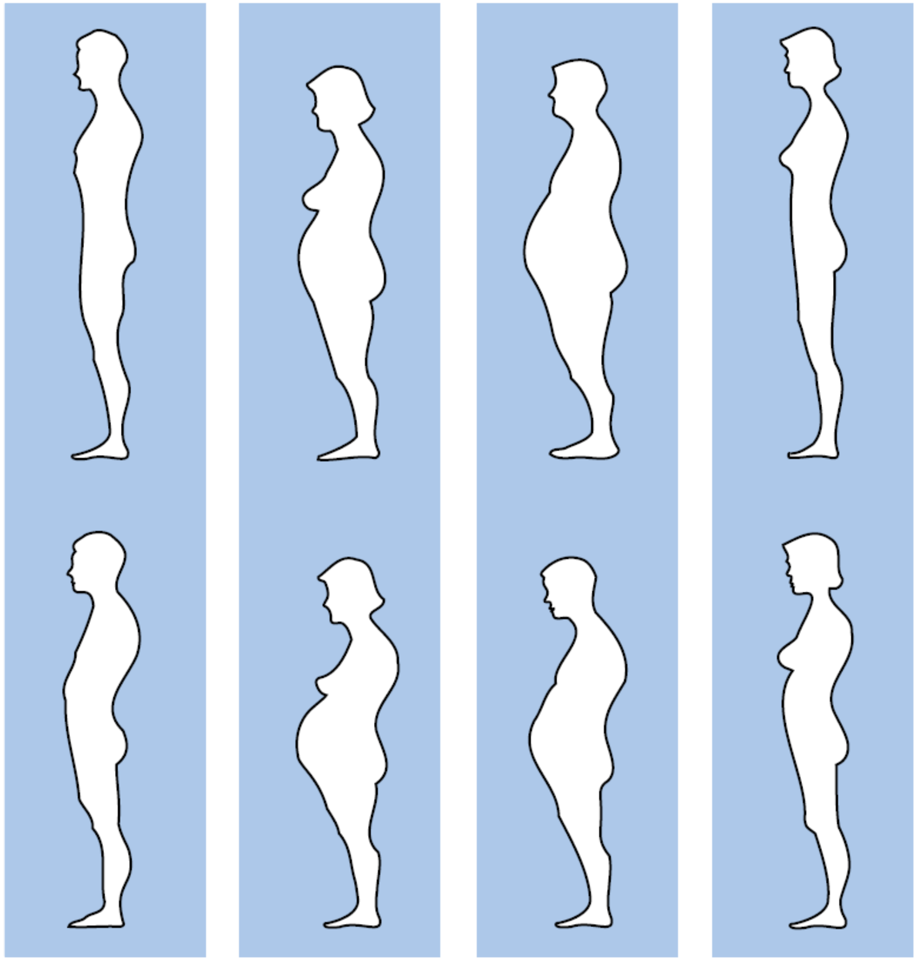


Figure 1b (p. 394)

Matched pairs design with four pairs of subjects. Separate random assignment of Drug 1
each pair.

Independent vs. Matched Pair: Expectation Each Pair to be Dependent Given the Subjects' Pre-existing Condition

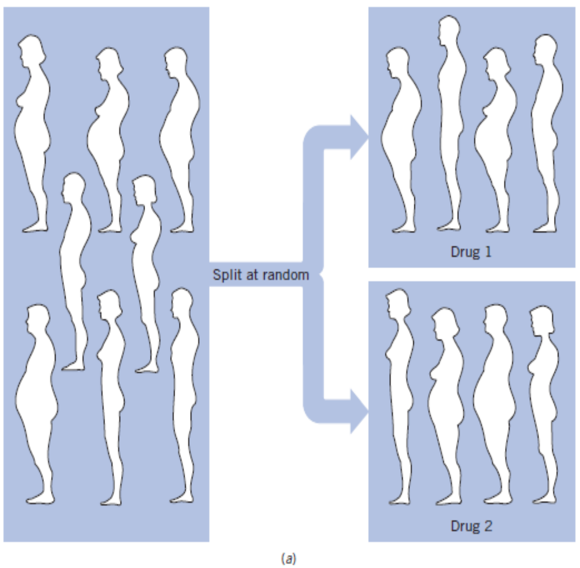


Figure 1a Independent samples, each of size 4.

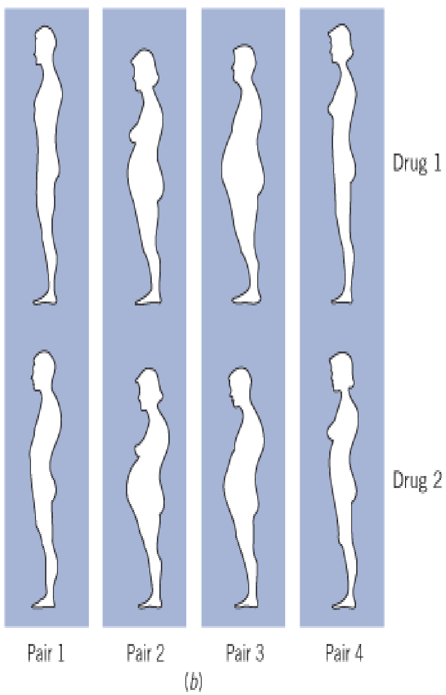


Figure 1b Matched pairs design with four pairs of subjects. Separate random assignment of Drug 1 each pair.

In contrast with the situation of Figure 1a, we would expect the responses of each pair to be dependent for the reason that they are governed by the same preexisting conditions of the subjects.

Study Design Determines Data Collection and the Key to Correct Analysis

In summary, a carefully planned experimental design is crucial to a successful comparative study. The design determines the structure of the data. In turn, the design provides the key to selecting an appropriate analysis.

10.4 Identify the following as either matched pair or independent samples. Also identify the experimental units, treatments, and response in each case.

(a) Twelve persons are given a high-potency vitamin C capsule once a day. Another twelve do not take extra vitamin C. Investigators record the number of colds in 5 winter months.

(b) One self-fertilized plant and one cross-fertilized plant are grown in each of 7 pots. Their heights are measured after 3 months.

(c) Ten newly married couples are interviewed. Both the husband and wife respond to the question, "How many children would you like to have?"

(d) Learning times are recorded for 5 dogs trained by a reward method and 3 dogs trained by a reward-punishment method.

Examples of Independent Samples vs. Matched Pair Samples

10.4

	Unit	Treatments	Response	Design
(a)	Person	vitamin C capsule no vitamin C	Number of colds in 5 months	Independent samples
(b)	Plant	self-fertilization cross-fertilization	Height after 3 months	Matched pair
(c)	Person	husband wife	Number of children	Matched pair
(d)	Dog	reward training reward punishment	Learning times	Independent samples

Inferences About Difference Between Means of Two Independent Samples

2. Inferences About the Difference of Means—Independent Large Samples

Here we discuss the methods of statistical inference for comparing two treatments or two populations on the basis of independent samples. Recall that with the **independent samples design**, a collection of $n_1 + n_2$ subjects is randomly divided into two groups and the responses are recorded. We conceptualize population 1 as the collection of responses that would result if a vast number of subjects were given treatment 1. Similarly, population 2 refers to the population of responses under treatment 2. The design of independent samples can then be viewed as one that produces unrelated random samples from two populations (see Figure 2). In other situations, the populations to be compared may be quite real entities. For instance, one may wish to compare the residential property values in the east suburb of a city to those in the west suburb. Here the issue of assigning experimental subjects to treatments does not arise. The collection of all residential properties in each suburb constitutes a population from which a sample will be drawn at random.

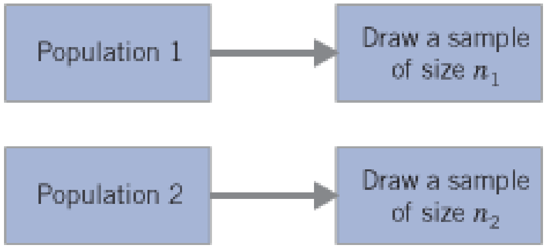


Figure 2 Independent random samples.

Design of Independent Samples

With the design of independent samples, we obtain

Sample	Summary Statistics	
$X_1, X_2, \ldots, X_{n_1}$ from population 1	$\overline{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$	$S_1^2 = \frac{\sum_{i=1}^{n_1} \left(X_i - \overline{X}\right)^2}{n_1 - 1}$
$Y_1, Y_2, \ldots, Y_{n_2}$ from population 2	$\overline{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$	$S_2^2 = \frac{\sum_{i=1}^{n_2} \left(Y_i - \overline{Y}\right)^2}{n_2 - 1}$

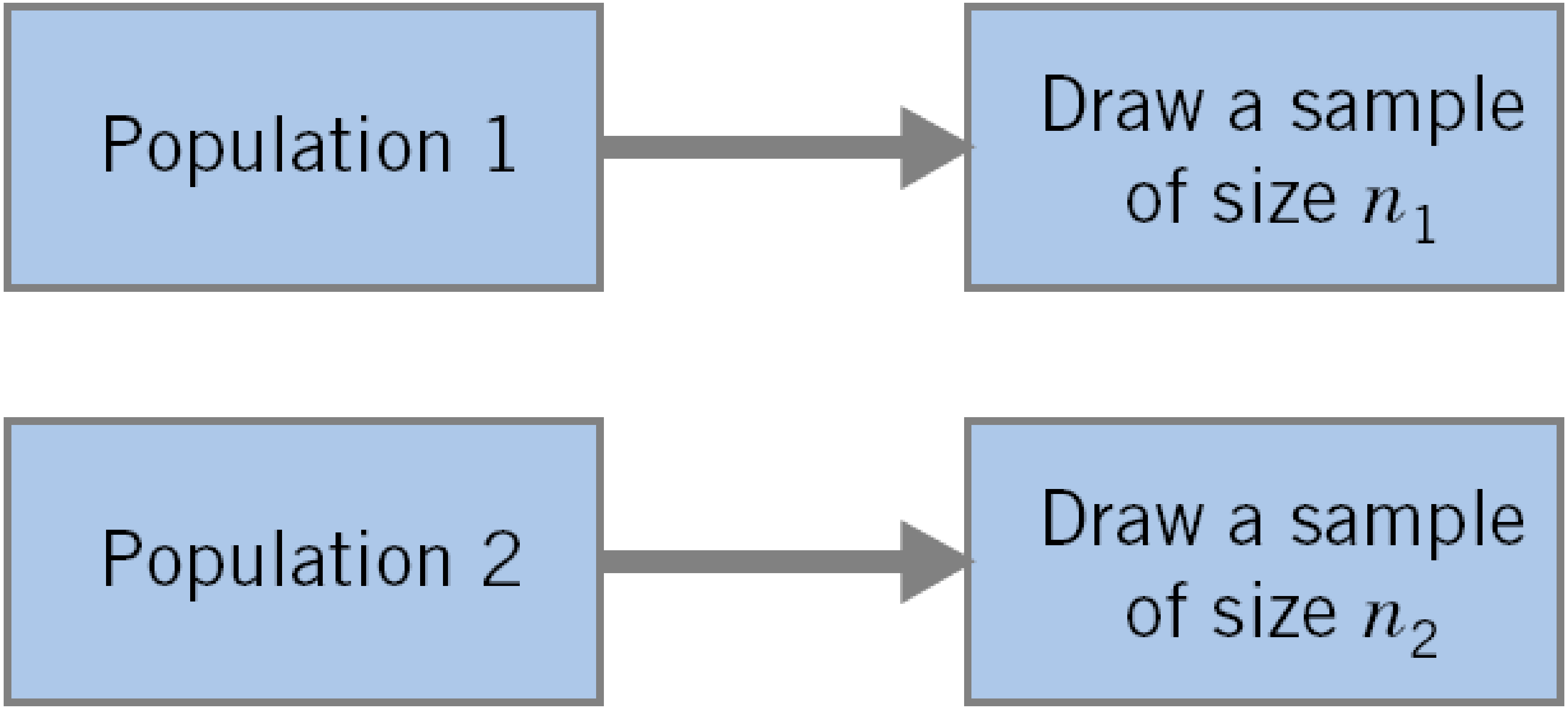


Figure 2 (p. 395)
Independent random samples.

Sample	Summary Statistics	
$X_1, X_2, \ldots, X_{n_1}$ from population 1	$\overline{X} = \dfrac{1}{n_1} \displaystyle\sum_{i=1}^{n_1} X_i$	$S_1^2 = \dfrac{\sum_{i=1}^{n_1} (X_i - \overline{X})^2}{n_1 - 1}$
$Y_1, Y_2, \ldots, Y_{n_2}$ from population 2	$\overline{Y} = \dfrac{1}{n_2} \displaystyle\sum_{i=1}^{n_2} Y_i$	$S_2^2 = \dfrac{\sum_{i=1}^{n_2} (Y_i - \overline{Y})^2}{n_2 - 1}$

Box 1 on Page 395

Conditions for Two Independent Random Samples

To make confidence statements or to test hypotheses, we specify a statistical model for the data.

Statistical Model: Independent Random Samples

1. $X_1, X_2, \ldots, X_{n_1}$ is a random sample of size n_1 from population 1 whose mean is denoted by μ_1 and standard deviation by σ_1.
2. $Y_1, Y_2, \ldots, Y_{n_2}$ is a random sample of size n_2 from population 2 whose mean is denoted by μ_2 and standard deviation by σ_2.
3. The samples are independent. In other words, the response measurements under one treatment are unrelated to the response measurements under the other treatment.

We now set our goal toward drawing a comparison between the mean responses of the two treatments or populations. In statistical language, we are interested in making inferences about the parameter

$$\mu_1 - \mu_2 = (\text{Mean of population } 1) - (\text{Mean of population } 2)$$

When the sample sizes are large, no additional assumptions are required.

Statistical Model: Independent Random Samples

1. $X_1, X_2, \ldots, X_{n_1}$ is a random sample of size n_1 from population 1 whose mean is denoted by μ_1 and standard deviation by σ_1.
2. $Y_1, Y_2, \ldots, Y_{n_2}$ is a random sample of size n_2 from population 2 whose mean is denoted by μ_2 and standard deviation by σ_2.
3. The samples are independent. In other words, the response measurements under one treatment are unrelated to the response measurements under the other treatment.

Box 2 on Page 396

Statistical Model: Independent Random Samples

Review: Constructing Confidence Intervals – 2 Independent Samples

Constructing Confidence Interval Two Independent Samples: Subtract the Means and Sum the Variance

2.1 ESTIMATION

Inferences about the difference $\mu_1 - \mu_2$ are naturally based on its estimate $\overline{X} - \overline{Y}$, the difference between the sample means. When both sample sizes n_1 and n_2 are large (say, greater than 30), $\overline{X}$ and $\overline{Y}$ are each approximately normal and their difference $\overline{X} - \overline{Y}$ is approximately normal with

Mean

$$E\left(\overline{X} - \overline{Y}\right) = \mu_1 - \mu_2$$

Variance

$$\text{Var}\left(\overline{X} - \overline{Y}\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Standard error

$$\text{S.E.}\left(\overline{X} - \overline{Y}\right) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Note: Because the entities $\overline{X}$ and $\overline{Y}$ vary in repeated sampling and independently of each other, the distance between them becomes more variable than the individual members. This explains the mathematical fact that the variance of the difference $\overline{X} - \overline{Y}$ equals the *sum* of the variances of $\overline{X}$ and $\overline{Y}$.

Z-Score for Confidence Interval of Two Independent Samples

When n_1 and n_2 are both large, the normal approximation remains valid if σ_1^2 and σ_2^2 are replaced by their estimators

$$S_1^2 = \frac{\sum_{i=1}^{n_1} (X_i - \overline{X})^2}{n_1 - 1} \quad \text{and} \quad S_2^2 = \frac{\sum_{i=1}^{n_2} (Y_i - \overline{Y})^2}{n_2 - 1}$$

We conclude that, when the sample sizes n_1 and n_2 are large,

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad \text{is approximately } N(0, 1)$$

Confidence Interval Formula for Difference Between Means of Two Independent Samples

A confidence interval for $\mu_1 - \mu_2$ is constructed from this sampling distribution. Similar to the single sample confidence interval, our confidence interval here is again of the form

$$\text{Estimate of parameter} \pm (z \text{ value}) \ (\text{estimated standard error})$$

Large Samples Confidence Interval for $\mu_1 - \mu_2$

When n_1 and n_2 are greater than 30, an approximate $100\,(1-\alpha)\,\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},\ \ \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ point of $N\,(0,\,1)$.

Large Samples Confidence Interval for $\mu_1 - \mu_2$

When n_1 and n_2 are greater than 30, an approximate $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \quad \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ point of $N(0, 1)$.

Box on Page 397

Large Samples Confidence Interval for $\mu_1 - \mu_2$

Example: Construct Confidence Interval for Difference in Mean Job Satisfaction Scores

Example 4

Large Samples Confidence Interval for Difference in Mean Job Satisfaction

A considerable proportion of a person's time is spent working, and satisfaction with the job and satisfaction with life tend to be related. Job satisfaction is typically measured on a four point scale

Very Dissatisfied	*A Little Dissatisfied*	*Moderately Satisfied*	*Very Satisfied*

A numerical scale is created by assigning 1 to very dissatisfied, 2 to a little dissatisfied, 3 to moderately satisfied, and 4 to very satisfied.
The responses of 226 firefighters and 247 office supervisors, presented in Exercise 10.12, yielded the summary statistics

	Fire-fighter	Office Supervisor
Mean	3.673	3.547
sd	.7235	.6089

Construct a 95% confidence interval for difference in mean job satisfaction scores.

	Fire-fighter	Office Supervisor
Mean	3.673	3.547
sd	0.7235	0.6089

Data on Page 398

Example Solution: Confidence Interval for Difference in Mean Job Satisfaction Scores

SOLUTION

Let μ_1 be the mean job satisfaction for firefighters and μ_2 the mean for office supervisors. We have

$$n_1 = 226 \quad \overline{x} = 3.673 \quad s_1 = .7235$$
$$n_2 = 247 \quad \overline{y} = 3.547 \quad s_2 = .6089$$
$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(.7235)^2}{226} + \frac{(.6089)^2}{247}} = .06018$$

For a 95% confidence interval, we use $z_{.025} = 1.96$ and calculate

$$\overline{x} - \overline{y} = 3.673 - 3.547 = .126$$
$$z_{.025}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.96 \times .06018 = .121$$

Therefore, a 95% confidence interval for $\mu_1 - \mu_2$ is given by

$$.126 \pm .121 \quad \text{or} \quad (.005, .247)$$

We are 95% confident that the mean score for firefighters is .005 units to .247 units higher than the mean score of office supervisors. This interval contains inconsequential positive values as well as positive values that are possibly important differences on the satisfaction scale.

Hypothesis Testing of Two Mean Differences: Use Z-Score as Test Statistic

2.2 HYPOTHESES TESTING

Let us turn our attention to testing hypotheses concerning $\mu_1 - \mu_2$. A test of the null hypothesis that the two population means are the same, $H_0: \mu_1 - \mu_2 = 0$, employs the test statistic

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

which is approximately $N(0, 1)$ when $\mu_1 - \mu_2 = 0$.

Example 5

Testing Equality of Mean Job Satisfaction

Do the data in Example 4 provide strong evidence that the mean job satisfaction of firefighters is different from the mean job satisfaction of office supervisors? Test at $\alpha = .02$.

SOLUTION

Because we are asked to show that the two means are different, we formulate the problem as testing

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{versus} \quad H_1: \mu_1 - \mu_2 \neq 0$$

Hypothesis Testing of Two Mean Differences: Use Z-Score as Test Statistic

SOLUTION

Because we are asked to show that the two means are different, we formulate the problem as testing

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{versus} \quad H_1: \mu_1 - \mu_2 \neq 0$$

We use the test statistic

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

and set a two-sided rejection region. Specifically, with $\alpha = .02$, we have $\alpha / 2 = .01$ and $z_{\alpha/2} = 2.33$, so the rejection region is $R: |Z| \geq 2.33$.

Using the sample data given in Example 4, we calculate

$$z = \frac{3.673 - 3.547}{\sqrt{\frac{(.7235)^2}{226} + \frac{(.6089)^2}{247}}} = \frac{.126}{.0618} = 2.04$$

Hypothesis Testing of Two Mean Differences: 2-Sided = or not =

Because the observed value $z = 2.04$ does not lie in the rejection region, we fail to reject the null hyphothesis at level $\alpha = .02$. The evidence against equal means is only moderately strong since the P-value is

$$P[Z < -2.04] + P[Z > 2.04] = .0207 + .0207 = .0414 \quad \text{(see Figure 3)}$$

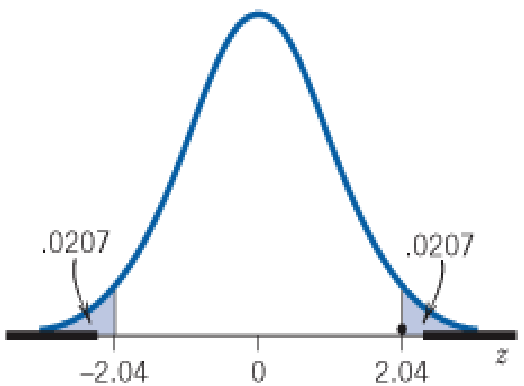


Figure 3 P-value with two-sided rejection region.

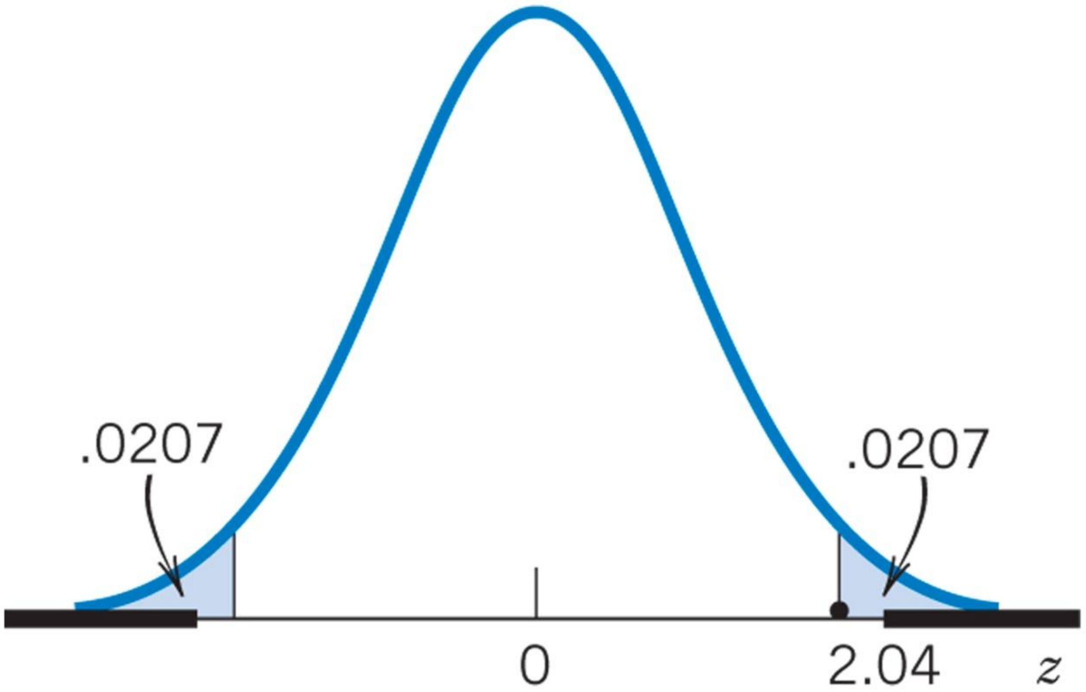


Figure 3 (p. 399)
P-value with two-sided rejection region.

Comparison of Chlorine Content in Two Lakes: One-Sided Hypothesis Test

Example 6

Large Samples Test with a One-Sided Alternative

In June two years ago, chemical analyses were made of 85 water samples (each of unit volume) taken from various parts of a city lake, and the measurements of chlorine content were recorded. During the next two winters, the use of road salt was substantially reduced in the catchment areas of the lake. This June, 110 water samples were analyzed and their chlorine contents recorded. Calculations of the mean and the standard deviation for the two sets of data give

	Chlorine Content	
	Two Years Ago	**Current Year**
Mean	18.3	17.8
Standard deviation	1.2	1.8

Test the claim that lower salt usage has reduced the amount of chlorine in the lake. Base your decision on the P-value.

	Chlorine Content	
	Two Years Ago	Current Year
Mean	18.3	17.8
Standard deviation	1.2	1.8

Example 6 data page 400

Chlorine content.

Comparison of Chlorine Content in Two Lakes: One-Sided Hypothesis Test or Lower Salt Reduced Chlorine i.e. >

Test the claim that lower salt usage has reduced the amount of chlorine in the lake. Base your decision on the P-value.

SOLUTION

Let μ_1 be the population mean two years ago and μ_2 the population mean in the current year. Because the claim is that μ_2 is less than μ_1, we formulate the hypotheses

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{versus} \quad H_1: \mu_1 - \mu_2 > 0$$

With the test statistic

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

the rejection region should be of the form $R: Z \geq c$ because H_1 is right-sided.

Comparison Two Samples: Use Z-Score

Using the data

$$n_1 = 85 \quad \overline{x} = 18.3 \quad s_1 = 1.2$$
$$n_2 = 110 \quad \overline{y} = 17.8 \quad s_2 = 1.8$$

we calculate

$$z = \frac{18.3 - 17.8}{\sqrt{\frac{(1.2)^2}{85} + \frac{(1.8)^2}{110}}} = \frac{.5}{.2154} = 2.32$$

The significance probability of this observed value is (see Figure 4)

$$P\text{–value} = P[Z \geq 2.32] = .0102$$

Z-Score Defines Rejection from Non-Rejection & Then Corresponding p-value

The significance probability of this observed value is (see Figure 4)

$$P\text{-value} = P[Z \geq 2.32] = .0102$$

Because the P-value is very small, we conclude that there is strong evidence to reject H_0 and support H_1.

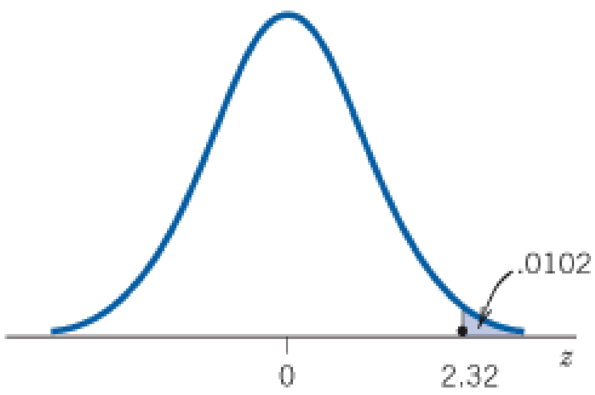


Figure 4 $P\text{-value} = .0102$.

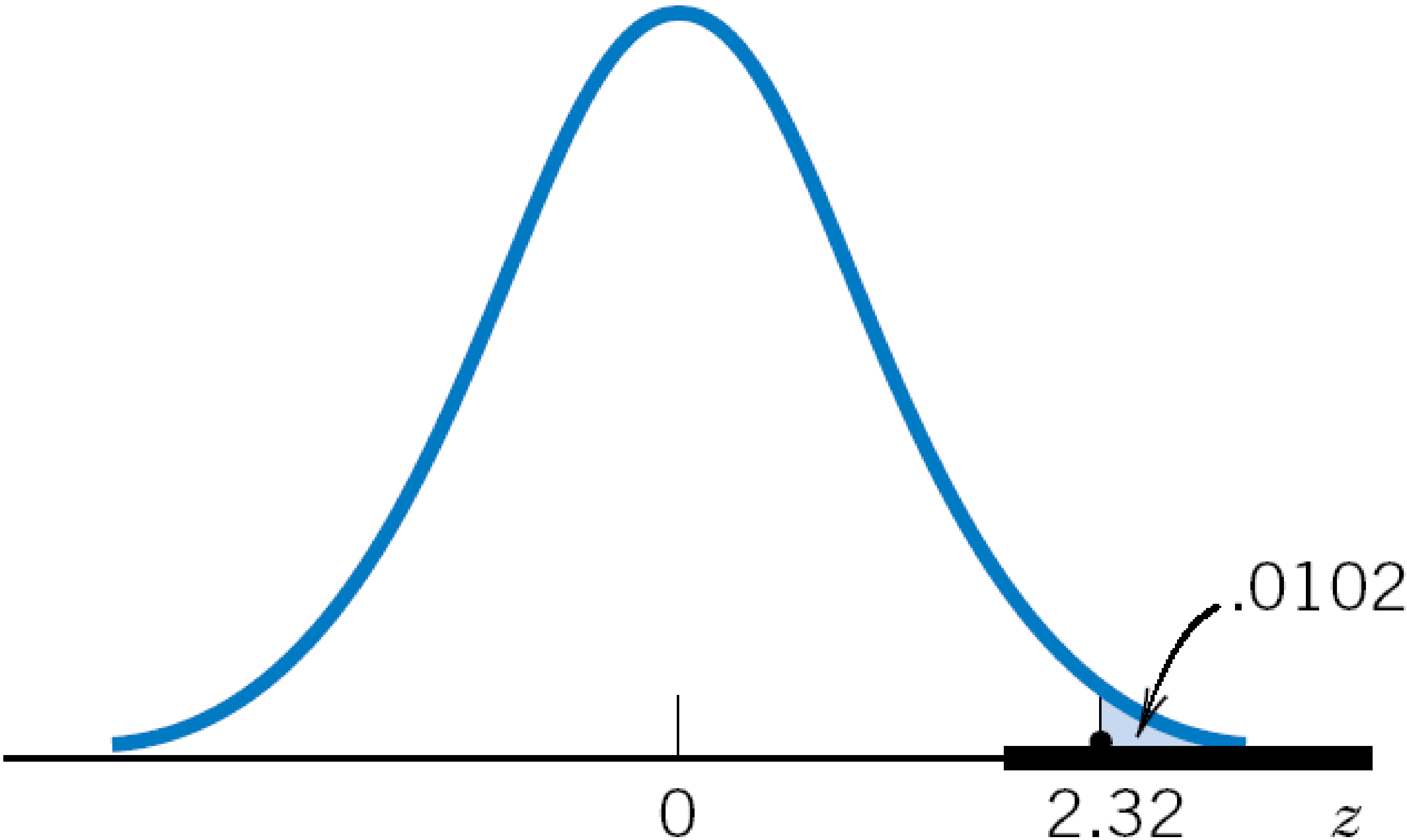


Figure 4 (p. 400)
P-value = .0102.

Hypothesis Testing of Two Sample Means from Two Large Samples n > 30

We summarize the procedure for testing $\mu_1 - \mu_2 = \delta_0$ where δ_0 is specified under the null hypothesis. The case $\mu_1 = \mu_2$ corresponds to $\delta_0 = 0$.

Testing $H_0: \mu_1 - \mu_2 = \delta_0$ with Large Samples

Test statistic:

$$Z = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Alternative Hypothesis	**Level α Rejection Region**
$H_1: \mu_1 - \mu_2 > \delta_0$	$R: Z \geq z_\alpha$
$H_1: \mu_1 - \mu_2 < \delta_0$	$R: Z \leq -z_\alpha$
$H_1: \mu_1 - \mu_2 \neq \delta_0$	$R: \lvert Z \rvert \geq z_{\alpha/2}$

Finally, we would like to emphasize that with large samples we can also learn about other differences between the two populations.

Testing $H_0: \mu_1 - \mu_2 = \delta_0$ with Large Samples

Test statistic:

$$Z = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

Alternative Hypothesis	Level α Rejection Region
$H_1: \mu_1 - \mu_2 > \delta_0$	$R: Z \geq z_\alpha$
$H_1: \mu_1 - \mu_2 < \delta_0$	$R: Z \leq -z_\alpha$
$H_1: \mu_1 - \mu_2 \neq \delta_0$	$R: \lvert Z \rvert \geq z_{\alpha/2}$

Boxes on Page 401

Radiance Intensity Levels Distinguishing Urban from Forest Area

Example 7

Large Samples Reveal Additional Differences between Populations

Natural resource managers have attempted to use the Satellite Landsat Multispectral Scanner data for improved landcover classification. When the satellite was flying over country known to consist of forest, the following intensities were recorded on the near-infrared band of a thermatic mapper. The sample has already been ordered.

77	77	78	78	81	81	82	82	82	82	82	83	83	84	84	84
84	85	86	86	86	86	86	87	87	87	87	87	87	87	89	89
89	89	89	89	89	90	90	90	91	91	91	91	91	91	91	91
91	91	93	93	93	93	93	93	94	94	94	94	94	94	94	94
94	94	94	94	95	95	95	95	95	96	96	96	96	96	96	97
97	97	97	97	97	97	97	97	98	99	100	100	100	100		
100	100	100	100	100	101	101	101	101	101	101	102				
102	102	102	102	102	103	103	104	104	104	105	107				

When the satellite was flying over urban areas, the intensities of reflected light on the same near-infrared band were

71	72	73	74	75	77	78	79	79	79	79	80	80	80	81	81	81
82	82	82	82	84	84	84	84	84	84	85	85	85	85	85	85	86
86	87	88	90	91	94											

If the means are different, the readings could be used to tell urban from forest area. Obtain a 95% confidence interval for the difference in mean radiance levels.

Confidence Interval for Mean Difference Between Forest and Urban Radiance Intensity Levels

SOLUTION

Computer calculations give

	Forest	Urban
Number	118	40
Mean	92.932	82.075
Standard deviation	6.9328	4.9789

and, for large sample sizes, the approximate 95% confidence interval for $\mu_1 - \mu_2$ is given by

$$\left(\overline{X} - \overline{Y} - z_{.025}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},\ \ \overline{X} - \overline{Y} + z_{.025}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$$

Confidence Interval for Mean Difference Between Forest and Urban Radiance Intensity Levels

Since $z_{.025} = 1.96$, the 95% confidence interval is calculated as

$$92.932 - 82.075 \pm 1.96\sqrt{\frac{(6.9328)^2}{118} + \frac{(4.9789)^2}{40}} \text{ or } (8.87,\ 12.84)$$

The mean for the forest is 8.87 to 12.84 levels of radiance higher than the mean for the urban areas. Because the sample sizes are large, we can also learn about other differences between the two populations. The stem-and-leaf displays and boxplots in Figure 5 reveal that there is some difference in the standard deviation as well as the means. The graphs further indicate a range of high readings that are more likely to come from forests than urban areas. This feature has proven helpful in discriminating between forest and urban areas on the basis of near-infrared readings.

Stem and Leaf and Boxplot to Illustrate Mean Difference Between Forest and Urban Radiance Intensity Levels

The mean for the forest is 8.87 to 12.84 levels of radiance higher than the mean for the urban areas. Because the sample sizes are large, we can also learn about other differences between the two populations. The stem-and-leaf displays and boxplots in Figure 5 reveal that there is some difference in the standard deviation as well as the means. The graphs further indicate a range of high readings that are more likely to come from forests than urban areas. This feature has proven helpful in discriminating between forest and urban areas on the basis of near-infrared readings.

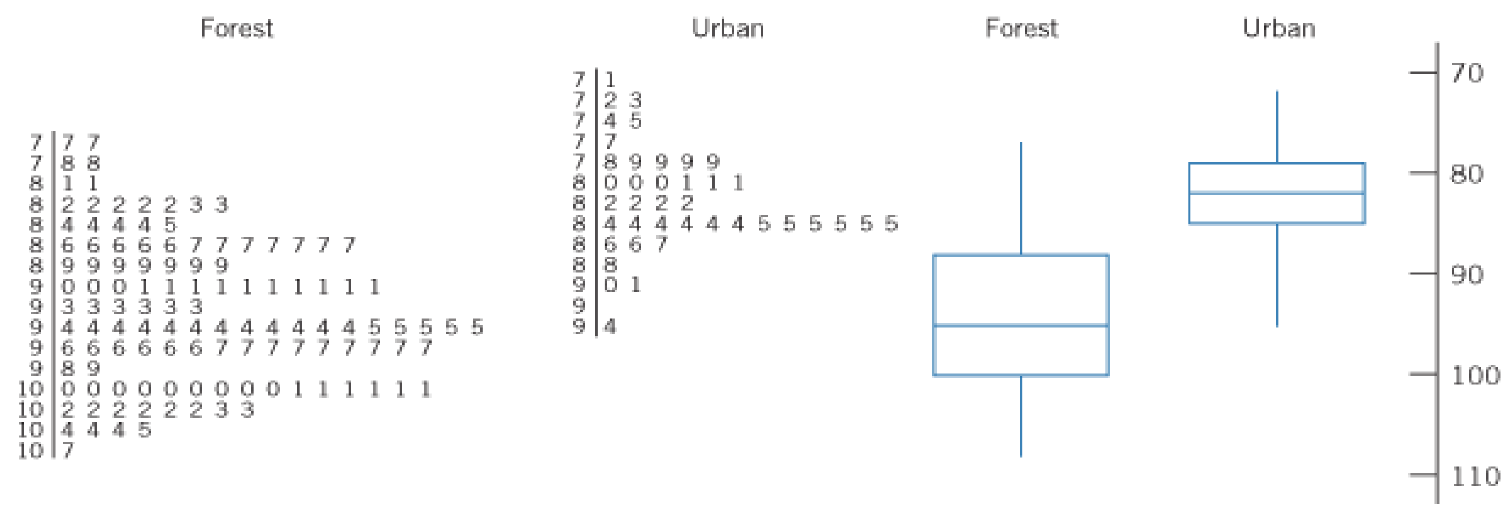


Figure 5 Stem-and-leaf displays and boxplots give additional information about population differences.

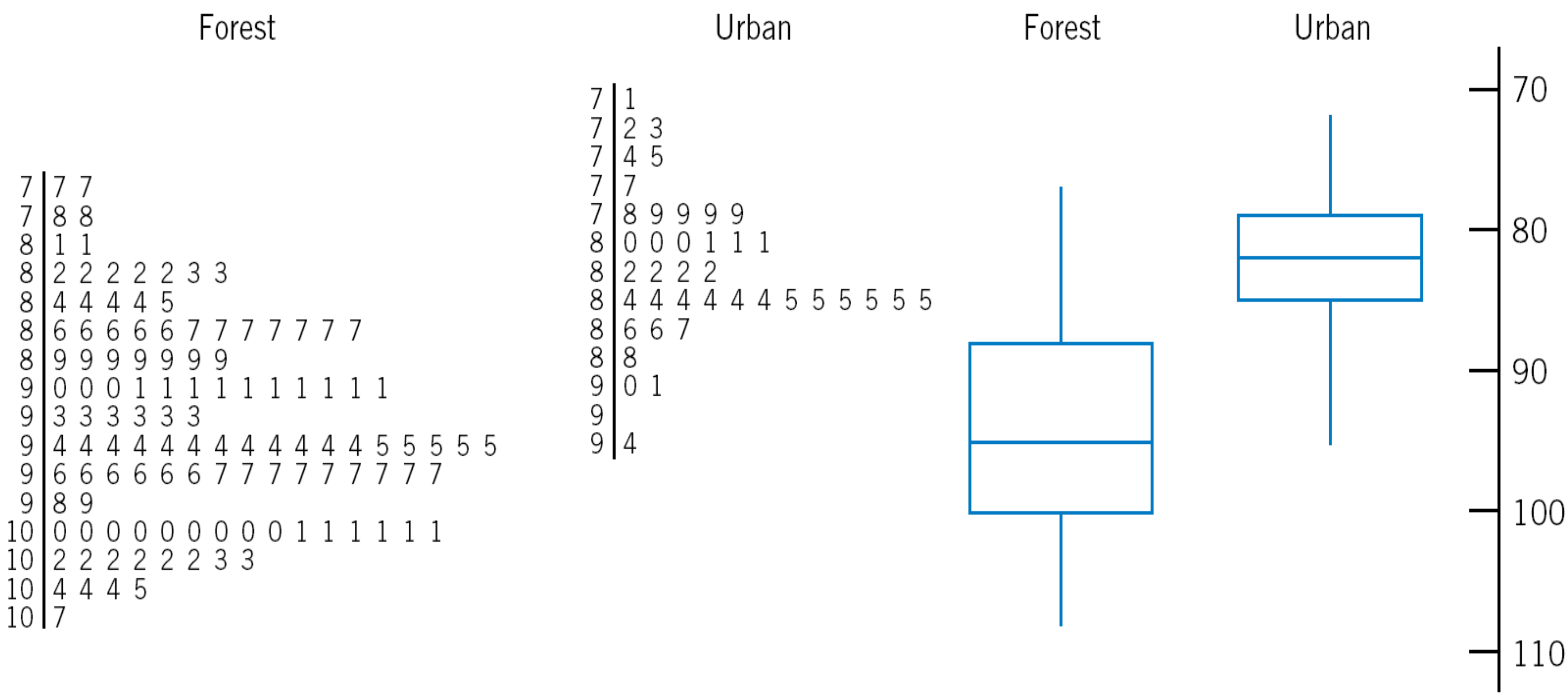


Figure 5 (p. 402)

Stem-and-leaf displays and boxplots give additional information about population differences.

Review: Comparing 2 Samples Variances or Standard Deviations Equal

Comparing Two Samples: Variances σ^2 or Standard Deviation σ Are Equal

3. Inferences About the Difference of Means—Independent Small Samples From Normal Populations

Not surprisingly, more distributional structure must be imposed before inference procedures can be developed for small samples. Here, we require that both population distributions be normal. A restriction to normal populations is not new and was already required for small sample inferences about a single mean. We treat the equal variance case first. It is the simplest.

3.1 INFERENCES: NORMAL POPULATIONS WITH EQUAL VARIANCES

The additional assumptions specify that the normal populations are normal and they also require the two standard deviations to be equal.

Additional Assumptions When the Sample Sizes Are Small

1. Both populations are normal.
2. The population standard deviations σ_1 and σ_2 are equal.

Additional Assumptions When the Sample Sizes Are Small

1. Both populations are normal.
2. The population standard deviations σ_1 and σ_2 are equal.

Box on Page 405

Small Sample Assumptions: Pool Variance and Then Divide by n_1 and n_2

The assumption requiring equal variability is somewhat artificial but we reserve comment until later. Letting σ denote the common standard deviation, we summarize.

Small Samples Assumptions

1. $X_1, X_2, \ldots, X_{n_1}$ is a random sample from $N(\mu_1, \sigma)$.
2. $Y_1, Y_2, \ldots, Y_{n_2}$ is a random sample from $N(\mu_2, \sigma)$.

   (*Note:* σ is the same for both distributions.)
3. $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are independent.

Estimation

Again, $\overline{X} - \overline{Y}$ is our choice for a statistic.

$$\text{Mean of } (\overline{X} - \overline{Y}) = E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2$$

$$\text{Var}(\overline{X} - \overline{Y}) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Pooled Variance or S^2: Numerator and Denominator Include Degrees of Freedom for Both Samples n_1 and n_2

The common variance σ^2 can be estimated by combining information provided by both samples. Specifically, the sum $\sum_{i=1}^{n_1} (X_i - \overline{X})^2$ incorporates $n_1 - 1$ pieces of information about σ^2, in view of the constraint that the deviations $X_i - \overline{X}$ sum to zero. Independently of this, $\sum_{i=1}^{n_2} (Y_i - \overline{Y})^2$ contains $n_2 - 1$ pieces of information about σ^2. These two quantities can then be combined,

$$\sum (X_i - \overline{X})^2 + \sum (Y_i - \overline{Y})^2$$

to obtain a pooled estimate of the common σ^2. The proper divisor is the sum of the component degrees of freedom, or $(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$.

Pooled Estimator of the Common σ^2

$$S^2_{\text{pooled}} = \frac{\sum_{i=1}^{n_1} (X_i - \overline{X})^2 + \sum_{i=1}^{n_2} (Y_i - \overline{Y})^2}{n_1 + n_2 - 2}$$

$$= \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}$$

Small Samples Assumptions

1. $X_1, X_2, \ldots, X_{n_1}$ is a random sample from $N(\mu_1, \sigma)$.
2. $Y_1, Y_2, \ldots, Y_{n_2}$ is a random sample from $N(\mu_2, \sigma)$.
   (*Note:* σ is the same for both distributions.)
3. $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are independent.

Pooled Estimator of the Common σ^2

$$S^2_{\text{pooled}} = \frac{\sum_{i=1}^{n_1} (X_i - \overline{X})^2 + \sum_{i=1}^{n_2} (Y_i - \overline{Y})^2}{n_1 + n_2 - 2}$$

$$= \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$

Boxes on Pages 405, 406

Small Samples Assumptions; Pooled Estimator of the Common σ^2

Small Samples Confidence Interval Calculation of Two Samples: Pooled Variance

Example 9

Calculating a Small Samples Confidence Interval

Beginning male and female accounting students were given a test and, on the basis of their answers, were assigned a computer anxiety score (CARS). Using the data given in Table D.4 of the Data Bank, obtain a 95% confidence interval for the difference in mean computer anxiety score between beginning male and female accounting students.

SOLUTION

The dot diagrams of these data, plotted in Figure 6, give the appearance of approximately equal amounts of variation.

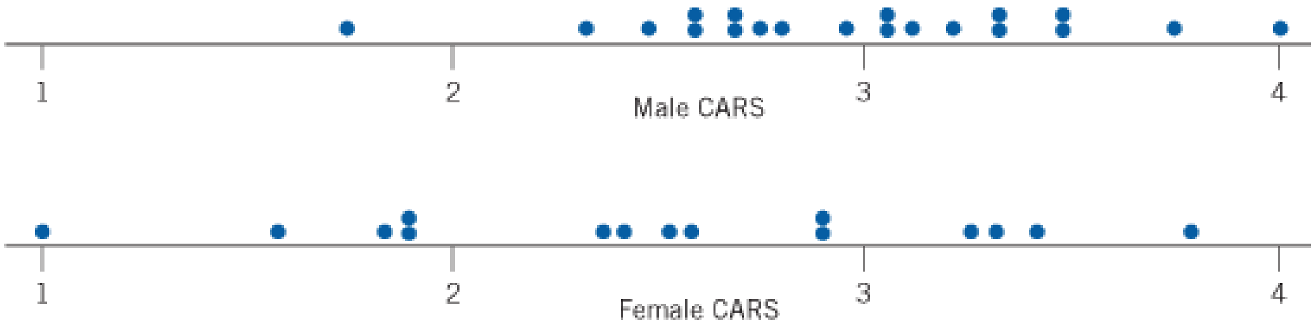


Figure 6 Dot diagrams of the computer anxiety data in Example 9.

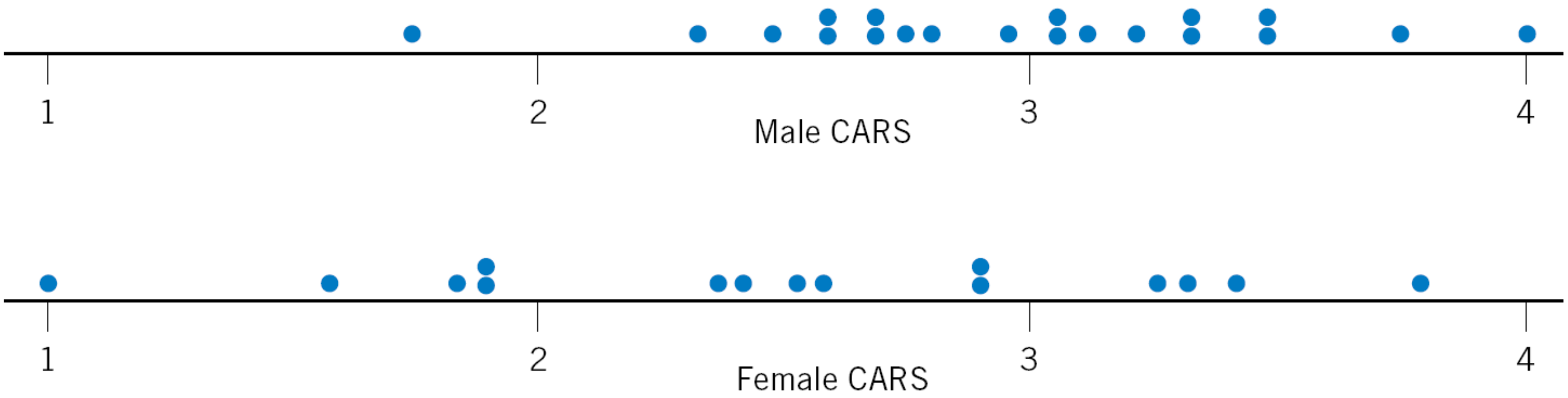


Figure 6 (p. 408)

Dot diagrams of the computer anxiety data in Example 9.

Confidence Interval for Two Samples: Calculate Mean and Standard Deviation Respectively for Each

We assume that the CARS data for both females and males are random samples from normal populations with means μ_1 and μ_2, respectively, and a common standard deviation σ. Computations from the data provide the summary statistics:

Female CARS

$$n_1 = 15 \quad \overline{x} = 2.514 \quad s_1 = .773$$

Male CARS

$$n_2 = 20 \quad \overline{y} = 2.963 \quad s_2 = .525$$

We calculate

$$s_{\text{pooled}} = \sqrt{\frac{14\,(.773)^2 + 19\,(.525)^2}{15 + 20 - 2}} = .642$$

Confidence Interval for Two Samples: Use t critical Value Multiplied by Pooled Standard Deviation of 2 Samples

With a 95% confidence interval $\alpha/2 = .025$ and consulting the t table, we find (interpolating) that $t_{.025} = 2.035$ for d.f. $= n_1 + n_2 - 2 = 33$. Thus a 95% confidence interval for $\mu_1 - \mu_2$ becomes

$$\bar{x} - \bar{y} \pm t_{0.25}\, s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 2.514 - 2.963 \pm 2.035 \times .642 \sqrt{\frac{1}{15} + \frac{1}{20}}$$

$$= -.449 \pm .446 \quad \text{or} \quad (-.895, -.003)$$

We can be 95% confident that the mean computer anxiety score for female beginning accounting students can be .003 to .895 units lower than the mean score for males.

This interval is quite wide. Certainly the values near $-.003$ represent technically insignificant differences.

Confidence Interval for $\mu_1 - \mu_2$ Small Samples

A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\overline{X} - \overline{Y} \pm t_{\alpha/2}\, S_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where

$$S^2_{\text{pooled}} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

and $t_{\alpha/2}$ is the upper $\alpha/2$ point of the t distribution with d.f. $= n_1 + n_2 - 2$.

Box on Page 407

Confidence Interval for μ_1 - μ_2.

Hypothesis Tests for Two Samples: Subtract the Means and Formulate Ho and H_1 or H_a

Tests of Hypotheses

Tests of hypotheses concerning the difference in means are based on a statistic having student's t distribution.

Testing H_0: $\mu_1 - \mu_2 = \delta_0$ with Small Samples and $\sigma_1 = \sigma_2$

Test statistic:

$$T = \frac{(\overline{X} - \overline{Y}) - \delta_0}{S_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{d.f.} = n_1 + n_2 - 2$$

Alternative Hypothesis	Level α Rejection Region
$H_1: \mu_1 - \mu_2 > \delta_0$	$R: T \geq t_\alpha$
$H_1: \mu_1 - \mu_2 < \delta_0$	$R: T \leq -t_\alpha$
$H_1: \mu_1 - \mu_2 \neq \delta_0$	$R: \lvert T \rvert \geq t_{\alpha/2}$

Testing H_0: $\mu_1 - \mu_2 = \delta_0$ with Small Samples and $\sigma_1 = \sigma_2$

Test statistic:

$$T = \frac{(\overline{X} - \overline{Y}) - \delta_0}{S_{\text{pooled}} \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad \text{d.f.} = n_1 + n_2 - 2$$

Alternative Hypothesis	Level α Rejection Region
$H_1 : \mu_1 - \mu_2 > \delta_0$	$R : T \ge t_\alpha$
$H_1 : \mu_1 - \mu_2 < \delta_0$	$R : T \le -t_\alpha$
$H_1 : \mu_1 - \mu_2 \ne \delta_0$	$R : \lvert T \rvert \ge t_{\alpha/2}$

Box on Page 409

Testing H_0: $\mu_1 - \mu_2 = \delta_0$ with small samples

Hypothesis Testing 2 Samples Equality of Means: Use t-Test With Pooled Square Root of Variance i.e. Standard Deviation

Example 10

Testing the Equality of Mean Computer Anxiety Scores

Refer to the computer anxiety scores (CARS) described in Example 9 and the summary statistics

Female CARS

$n_1 = 15 \quad \overline{x} = 2.514 \quad s_1 = .773$

Male CARS

$n_2 = 20 \quad \overline{y} = 2.963 \quad s_2 = .525$

Do these data strongly indicate that the mean score for females is lower than that for males? Test at level $\alpha = .05$.

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom.

Box on Page 407

Hypothesis Testing 2 Samples t-Test With Standard Deviation: To Find t Critical Value Must Pool Degrees of Freedom

SOLUTION

We are seeking strong evidence in support of the hypothesis that the mean computer anxiety score for females (μ_1) is less than the mean score for males. Therefore the alternative hypothesis should be taken as $H_1: \mu_1 < \mu_2$ or $H_1: \mu_1 - \mu_2 < 0$, and our problem can be stated as testing

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{versus} \quad H_1: \mu_1 - \mu_2 < 0$$

We employ the test statistic

$$T = \frac{\overline{X} - \overline{Y}}{S_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{d.f.} = n_1 + n_2 - 2$$

and set the left-sided rejection region $R: T \leq -t_{.05}$. For $\text{d.f.} = n_1 + n_2 - 2 = 33$, we approximate the tabled value as $t_{.05} = 1.692$. The rejection region is $R: T \leq -1.692$ as illustrated in Figure 7.

Right-Sided Hypothesis Test of Mean Anxiety Score Females Less Than Males: Compare Calculated t to t Critical Value

and set the left-sided rejection region $R: T \leq -t_{.05}$. For d.f. $= n_1 + n_2 - 2 = 33$, we approximate the tabled value as $t_{.05} = 1.692$. The rejection region is $R: T \leq -1.692$ as illustrated in Figure 7.

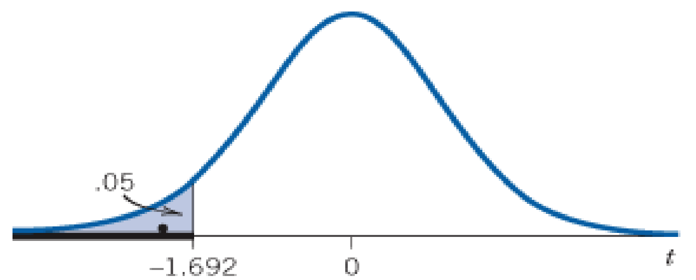


Figure 7 Rejection region for Example 10.

With $S_{\text{pooled}} = .642$ already calculated in Example 9, the observed value of the test statistic T is

$$t = \frac{2.514 - 2.963}{.642\sqrt{\frac{1}{15} + \frac{1}{20}}} = \frac{-.449}{.2193} = -2.05$$

This value lies in the rejection region R. Consequently, at the .05 level of significance, we reject the null hypothesis in favor of the alternative hypothesis that males have a higher mean computer anxiety score.

A computer calculation gives a P-value of about .025 so the evidence of H_1 is moderately strong.

Confidence Interval for 2 Samples With Unequal Variances: Using Pooled σ

Example 11

A Confidence Interval Shows Mothers of Males Have Milk with Higher Energy Content

Researchers, interested in the energy content of mothers' milk, conducted a study with 25 well-nourished, healthy mothers exclusively breast feeding their 2- to 5-month old babies. All of the participants lived in Massachusetts.[3] C. E. Powe, C. D. Knott, and N. Conklin-Brittain, "*Infant sex predicts breast milk energy content*," *American Journal of Human Biology* 22 (2010), pp. 50-54. Thirteen of the mothers had sons and twelve had daughters. The summary statistics

Energy (kcal/100 ml)

	Males	**Females**
Number	13	12
Mean	75.56	60.81
Standard deviation	19.37	15.64

are obtained in the study.
Does the gender of baby effect the amount of energy available in the mothers' milk? If so, what can you say about the mean difference?

Base your answers on a 95% confidence interval for the difference of mean energy density between the mothers' milk of male babies and and the mothers' milk of female babies.

Confidence Interval for 2 Samples With Unequal Variances: Using Pooled σ

SOLUTION

Since the two standard deviations are not very different, we use the pooled estimate of variance

$$s_{\text{pooled}} = \sqrt{\frac{13\,(19.37)^2 + 12\,(15.64)^2}{13 + 12 - 2}} = 17.684$$

For a 95% confidence interval, $\alpha/2 = .025$, and we find $t_{.025} = 2.069$ for d.f. $= 23$. The resulting 95% confidence interval is

$$\bar{x} - \bar{y} \;\pm\; t_{.025}\, s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 75.56 - 60.81 \pm 2.069 \times 17.684 \sqrt{\frac{1}{13} + \frac{1}{12}}$$

$$= 14.75 \pm 14.647$$

or $.10 < \mu_1 - \mu_2 < 29.40$

Confidence Interval for 2 Samples With Unequal Standard Deviation: Using Pooled σ

The 95% confidence interval includes only positive values for $\mu_1 - \mu_2$. With 95% confidence, we conclude that the mean energy density is larger when the baby is male than when it is female. However, the mean difference could be almost as small as .10 kcal/100ml which is likely not scientifically significant. It also could be almost as large as 29.4 kcal/100ml which is an important difference. A larger study is required to more accurately assess the mean difference.

This study involved well nourished healthy mothers in Massachusetts but they were not a random sample of new mothers. Generalizations are difficult, especially extrapolations to under-nourished mothers.

Pooling Variance vs. Not Pooling Standard Deviation When Constructing Confidence Intervals

Deciding Whether or not to Pool

Our preceding discussion of large and small sample inferences raises a few questions:

For small sample inference, why do we assume the population standard deviations to be equal when no such assumption was needed in the large samples case?

When should we be wary about this assumption, and what procedure should we use when the assumption is n
reasonable?

Learning statistics would be a step simpler if the ratio

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

had a t distribution for small samples from normal populations. Unfortunately, statistical theory proves it otherwise. The distribution of this ratio is *not* a t and, worse yet, it depends on the unknown quantity σ_1/σ_2.

Pooled Standard Deviation Rule of Thumb: If S_1/S_2 smaller than ½ or Greater than 2, Do NOT Pool

The assumption $\sigma_1 = \sigma_2$ and the change of the denominator to $S_{\text{pooled}}\sqrt{1/n_1 + 1/n_2}$ allow the t-based inferences to be valid. However, the $\sigma_1 = \sigma_2$ restriction and accompanying pooling are not needed in large samples where a normal approximation holds.

With regard to the second question, the relative magnitude of the two sample standard deviations s_1 and s_2 should be the prime consideration. The assumption $\sigma_1 = \sigma_2$ is reasonable if s_1/s_2 is not very much different from 1. As a working rule, the range of values $\frac{1}{2} \le s_1 \big/ s_2 \le 2$ may be taken as reasonable cases for making the assumption $\sigma_1 = \sigma_2$ and hence for pooling. If s_1/s_2 is seen to be smaller than $\frac{1}{2}$ or larger than 2, the assumption $\sigma_1 = \sigma_2$ would be suspect. In that case, some approximate methods of inference about $\mu_1 - \mu_2$ are available.

Review: Two Samples With Unequal Variances or Standard Deviations

Two Samples with Unequal Variances or Standard Deviations i.e. σ

3.2 INFERENCES: NORMAL POPULATIONS WITH UNEQUAL VARIANCES

When the sample sizes are not large and the observed standard deviations quite different, we can still proceed to test hypotheses and obtain confidence intervals provided that the populations are normal. The procedure we introduce is somewhat complex but is widely used in most statistical software programs. Consequently, using software is the preferred approach.

Inference procedures are developed based on a statistic that is the same as the large sample statistic for comparing two means. However, we use the notation T^*, where

$$T^* = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Because the sample sizes are small, we chose to approximate the distribution of T^* by a t distribution. But, its degrees of freedom are estimated using the sample standard deviations. As a consequence, the distribution of T^* is only approximately a t distribution.

Estimating Degrees of Freedom for Two Samples with Unequal Standard Deviation

This approach of approximating the distribution of T^* by a t, is sometimes called the **Satterthwaite correction** after the person who derived the approximation.

Normal Populations with $\sigma_1 \neq \sigma_2$

Let $\delta = \mu_1 - \mu_2$. When the sample sizes n_1 and n_2 are not large and $\sigma_1 \neq \sigma_2$

$$T^* = \frac{\overline{X} - \overline{Y} - \delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

is approximately distributed as a t with estimated degrees of freedom.

The estimated degrees of freedom depend on the observed values of the sample variances s_1^2 and s_2^2.

$$\text{estimated degrees of freedom} = \nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

The estimated degrees of freedom ν are often rounded down to an integer so a t table can be consulted.

Normal Populations with $\sigma_1 \neq \sigma_2$

Let $\delta = \mu_1 - \mu_2$. When the sample sizes n_1 and n_2 are not large and $\sigma_1 \neq \sigma_2$

$$T^* = \frac{\overline{X} - \overline{Y} - \delta}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

is approximately distributed as a t with estimated degrees of freedom.

The estimated degrees of freedom depend on the observed values of the sample variances s_1^2 and s_2^2.

estimated degrees of freedom =

$$v = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{s_1^2}{n_1}\right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{s_2^2}{n_2}\right)^2}$$

Box on Page 413

Small Sample Inferences for μ_1 - μ_2 When the Populations Are Normal But σ_1 or σ_2 Are Not Assumed to Be Equal

Hypothesis Testing and Confidence Interval For Two Sample Means Using Pooled Standard Deviation for Unequal Variances

Example 12

Testing Equality and Confidence Intervals Concerning Mean of Green Gas

One process of making green gasoline, not just a gasoline additive, takes biomass in the form of sucrose and converts it into gasoline using catalytic reactions. This research is still at the pilot plant stage. At one step in a pilot plant process, the product volume (liters) consists of carbon chains of length 3. Nine runs were made with each of two catalysts and the product volumes measured.

catalyst 1	1.86	2.05	2.06	1.88	1.75	1.64	1.86	1.75	2.13
catalyst 2	.32	1.32	.93	.84	.55	.84	.37	.52	.34

The sample sizes $n_1 = n_2 = 9$ and the summary statistics are

$$\bar{x} = 1.887, \quad s_1^2 = .0269 \quad \bar{y} = .670 \quad s_2^2 = .1133$$

(a) Is the mean yield with catalyst 1 more than .80 liters higher than the yield with catalyst 2? Test with $\alpha = 0.05$.

(b) Find a 98% confidence interval for the difference of means.

Hypothesis Testing For Two Sample Means Using Pooled Standard Deviation Formula for Unequal Variances

SOLUTION

(a) The test concerns the $\delta = \mu_1 - \mu_2$ and we wish to show that δ is greater than .80. Therefore we formulate the testing problem as

$$H_0: \mu_1 - \mu_2 = .80 \quad \text{versus} \quad H_1: \mu_1 - \mu_2 > .80$$

The sample sizes $n_1 = n_2 = 9$ are small and there are no outliers or obvious departures from normality

However, $s_2^2 = .1133$ is more than four times $s_1^2 = .0269$. We should not pool.

To apply the approximate t, we first estimate the degrees of freedom

$$\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{S_2^2}{n_2}\right)^2} = \frac{\left(\frac{.0269}{9} + \frac{.1133}{9}\right)^2}{\frac{1}{9-1}\left(\frac{.0269}{9}\right)^2 + \frac{1}{9-1}\left(\frac{.1133}{9}\right)^2} = 11.60$$

Confidence Interval For Two Sample Means Using Pooled Standard Deviation Formula for Unequal Variances

To use the t table, we round down to 11 and then obtain $t_{0.025} = 2.201$ for 11 degrees of freedom.

The rejection region becomes

$$T^* = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \geq 2.201$$

The calculated value of the test statistic is

$$t^* = \frac{1.887 - .670 - .80}{\sqrt{\frac{.0269}{9} + \frac{.1133}{9}}} = \frac{.417}{.1248} = 3.34$$

which falls in the rejection region. So, we reject the null hypothesis at $\alpha = .05$ and conclude that the mean product volume from catalyst 1 is more than .80 liters higher.

(b) Since $t_{.01} = 2.718$, the approximate 98% confidence interval is

$$\overline{x} - \overline{y} \pm 2.718\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.887 - .670 \pm 2.718\ (.1248) \text{ or } (.88,\ 1.56)$$

Review: Randomization in Research to Reduce Bias

Randomization in Research: No Particular Selection Method- Chance Assignment Reduces Bias

4. Randomization and Its Role in Inference

We have presented the methods of drawing inferences about the difference between two population means. Let us now turn to some important questions regarding the design of the experiment or data collection procedure. The manner in which experimental subjects are chosen for the two treatment groups can be crucial. For example, suppose that a remedial-reading instructor has developed a new teaching technique and is permitted to use the new method to instruct half the pupils in the class. The instructor might choose the most alert or the students who are more promising in some other way, leaving the weaker students to be taught in the conventional manner. Clearly, a comparison between the reading achievements of these two groups would not just be a comparison of two teaching methods. A similar fallacy can result in comparing the nutritional quality of a new lunch package if the new diet is given to a group of children suffering from malnutrition and the conventional diet is given to a group of children who are already in good health.

When the assignment of treatments to experimental units is under our control, steps can be taken to ensure a valid comparison between the two treatments. At the core lies the principle of impartial selection, or **randomization**. The choice of the experimental units for one treatment or the other must be made by a chance mechanism that does not favor one particular selection over any other. It must not be left to the discretion of the experimenters because, even unconsciously, they may be partial to one treatment.

Randomization Procedure for Comparing Two Treatments

Suppose that a comparative experiment is to be run with N experimental units, of which n_1 units are to be assigned to treatment 1 and the remaining $n_2 = N - n_1$ units are to be assigned to treatment 2. The principle of randomization tells us that the n_1 units for treatment 1 must be chosen at random from the available collection of N units—that is, in a manner such that all $\binom{N}{n_1}$ possible choices are equally likely to be selected.

Randomization Procedure for Comparing Two Treatments

From the available $N = n_1 + n_2$ experimental units, choose n_1 units at random to receive treatment 1 and assign the remaining n_2 units to treatment 2. The random choice entails that all $\binom{N}{n_1}$ possible selections are equally likely to be chosen.

As a practical method of random selection, we can label the available units from 1 to N. Then read random digits from Table 1, Appendix B, until n_1 different numbers between 1 and N are obtained. These n_1 experimental units receive treatment 1 and the remaining units receive treatment 2. For a quicker and more efficient means of random sampling, one can use the computer (see, for instance, the Technology section of Chapter 4).

Prospective Data Collection Randomization vs. Retrospective Observational Data Analysis

Although randomization is not a difficult concept, it is one of the most fundamental principles of a good experimental design. It guarantees that uncontrolled sources of variation have the same chance of favoring the response of treatment 1 as they do of favoring the response of treatment 2. Any systematic effects of uncontrolled variables, such as age, strength, resistance, or intelligence, are chopped up or confused in their attempt to influence the treatment responses.

Randomization prevents uncontrolled sources of variation from influencing the responses in a systematic manner.

Of course, in many cases, the investigator does not have the luxury of randomization. Consider comparing crime rates of cities before and after a new law. Aside from a package of criminal laws, other factors such as poverty, inflation, and unemployment play a significant role in the prevalence of crime. As long as these contingent factors cannot be regulated during the observation period, caution should be exercised in crediting the new law if a decline in the crime rate is observed or discrediting the new law if an increase in the crime rate is observed. When randomization cannot be performed, extreme caution must be exercised in crediting an apparent difference in means to a difference in treatments. The differences may well be due to another factor.

Randomization Procedure for Comparing Two Treatments

From the available $N = n_1 + n_2$ experimental units, choose n_1 units at random to receive treatment 1 and assign the remaining n_2 units to treatment 2. The random choice entails that all $\binom{N}{n_1}$ possible selections are equally likely to be chosen.

Randomization prevents uncontrolled sources of variation from influencing the responses in a systematic manner.

Boxes on Page 417

Randomization Procedure for Comparing Two Treatments

Matched Pair Comparisons Study Design: Used to Ensure Treatment Efficacy or Effectiveness

5. Matched Pairs Comparisons

In comparing two treatments, it is desirable that the experimental units or subjects be as alike as possible, so that a difference in responses between the two groups can be attributed to differences in treatments. If some identifiable conditions vary over the units in an uncontrolled manner, they could introduce a large variability in the measurements. In turn, this could obscure a real difference in treatment effects. On the other hand, the requirement that all subjects be alike may impose a severe limitation on the number of subjects available for a comparative experiment. To compare two analgesics, for example, it would be impractical to look for a sizable number of patients who are of the same sex, age, and general health condition and who have the same severity of pain. Aside from the question of practicality, we would rarely want to confine a comparison to such a narrow group. A broader scope of inference can be attained by applying the treatments on a variety of patients of both sexes and different age groups and health conditions.

Matching or Blocking Fundamental Element of Matched Pairs Study Design

The concept of **matching** or **blocking** is fundamental to providing a compromise between the two conflicting requirements that the experimental units be alike and also of different kinds. The procedure consists of choosing units in pairs or blocks so that the units in each block are similar and those in different blocks are dissimilar. One of the units in each block is assigned to treatment 1, the other to treatment 2. This process preserves the effectiveness of a comparison within each block and permits a diversity of conditions to exist in different blocks. Of course, the treatments must be allotted to each pair randomly to avoid selection bias. This design is called a **matched pairs design** or **sampling**. For example, in studying how two different environments influence the learning capacities of preschoolers, it is desirable to remove the effect of heredity: Ideally, this is accomplished by working with twins.

Matched Pairs Design

Matched pair	Experimental units	
1	2	1
2	1	2
3	1	2
⋮	⋮	⋮
n	2	1

Units in each pair are alike, whereas units in different pairs may be dissimilar. In each pair, a unit is chosen at random to receive treatment 1, the other unit receives treatment 2.

Matched Pairs Design

Matched pair	Experimental units	
1	2	1
2	1	2
3	1	2
⋮	⋮	⋮
n	2	1

Units in each pair are alike, whereas units in different pairs may be dissimilar. In each pair, a unit is chosen at random to receive treatment 1, the other unit receives treatment 2.

Box on Page 419

Matched Pairs Design

Detecting Differences in Matched Pair Study Design: Use Paired-Difference Formula

The structure of the observations in a paired comparison is given below, where X and Y denote the responses to treatments 1 and 2, respectively. The difference between the responses in each pair is recorded in the last column, and the summary statistics are also presented.

Structure of Data for a Matched Pair Comparison

Pair	Treatment 1	Treatment 2	Difference
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_2	Y_2	$D_2 = X_2 - Y_2$
.	.	.	.
.	.	.	.
.	.	.	.
n	X_n	Y_n	$D_n = X_n = Y_n$

The differences $D_1, D_2, \ldots, D_n$ are a random sample.

Summary statistics:

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \quad S_D^2 = \frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1}$$

Basic Assumption of Paired-Difference: If Mean Difference Equals Zero Then The Two Treatments are Equivalent

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \quad S_D^2 = \frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1}$$

Although the pairs (X_i, Y_i) are independent of one another, X_i and Y_i within the ith pair will usually be dependent. In fact, if the pairing of experimental units is effective, we would expect X_i and Y_i to be relatively large or small together. Expressed in another way, we would expect (X_i, Y_i) to have a high positive correlation. Because the differences $D_i = X_i - Y_i$, $i = 1, 2, \ldots, n$ are freed from the block effects, it is reasonable to assume that they constitute a random sample from a population with mean μ_D and variance σ_D^2, where μ_D represents the mean difference of the treatment effects. In other words,

$$E(D_i) = \mu_D \quad \text{and} \quad \operatorname{Var}(D_i) = \sigma_D^2 \text{ for } i = 1, \ldots, n$$

If the mean difference μ_D is zero, then the two treatments can be considered equivalent. A positive μ_D signifies that treatment 1 has a higher mean response than treatment 2. Considering $D_1, \ldots, D_n$ to be a single random sample from a population, we can immediately apply the techniques discussed in Chapters 8 and 9 to learn about the population mean μ_D.

Pairing (or Blocking)

Pairing like experimental units according to some identifiable characteristic(s) serves to remove this source of variation from the experiment.

Structure of Data for a Matched Pair Comparison

Pair	Treatment 1	Treatment 2	Difference
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_2	Y_2	$D_2 = X_2 - Y_2$
.	.	.	.
.	.	.	.
.	.	.	.
n	X_n	Y_n	$D_n = X_n - Y_n$

The differences $D_1, D_2, \ldots, D_n$ are a random sample.
Summary statistics:

$$\overline{D} = \frac{1}{n} \sum_{i=1}^{n} D_i \qquad S_D^2 = \frac{\sum_{i=1}^{n} (D_i - \overline{D})^2}{n - 1}$$

Boxes on Page 420

Pairing (or Blocking; Structure of Data for a Matched Pair Comparison

Statistics, 7/E by Johnson and Bhattacharyya

Review: Matched Pair Differences – Confidence Intervals

Matched Pair Difference Confidence Interval and Hypothesis Testing

5.1 INFERENCES FROM A LARGE NUMBER OF MATCHED PAIRS

As we learned in Chapter 8, the assumption of an underlying normal distribution can be relaxed when the sample size is large. The central limit theorem applied to the differences $D_1, \ldots, D_n$ suggests that when n is large, say, greater than 30,

$$\frac{\overline{D} - \mu_D}{S_D / \sqrt{n}} \quad \text{is approximately } N(0, 1)$$

Inferences can then be based on the percentage points of the $N(0, 1)$ distribution or, equivalently, those of the t distribution, with the degrees of freedom marked "infinity."

Matched Pairs

Twin Studies: Test Nature/Nurture Hypotheses or Influence of Environment vs. Genetics

digitalskillet / iStockphoto

Identical twins are the epitome of matched pair experimental subjects. They are matched with respect to not only age but also a multitude of genetic factors. Social scientists, trying to determine the influence of environment and heredity, have been especially interested in studying identical twins who were raised apart. Observed differences in IQ and behavior are then supposedly due to environmental factors.

When the subjects are animals like mice, two from the same litter can be paired. Going one step further, genetic engineers can now provide two identical plants or small animals by cloning these subjects.

Paired-Differences Confidence Interval and Hypothesis Test For Large Sample

Example 13

Does Conditioning Reduce Percent Body Fat?

A conditioning class is designed to introduce students to a variety of training techniques to improve fitness and flexibility. The percent body fat was measured at the start of the class and at the end of the semester. The data for 81 students are given in Table D.5 of the Data Bank.

(a) Obtain a 98% confidence interval for the mean reduction in percent body fat.

(b) Test, at $\alpha = .01$, to establish the claim that the conditioning class reduces the mean percent body fat.

SOLUTION

(a) Each subject represents a block that produces one measurement of percent body fat at the start of the semester (x) and one at the end (y). The 81 paired differences $d_i = x_i - y_i$ are summarized using a computer.

	N	Mean	StDev	SE Mean
Difference	81	3.322	2.728	0.303

Paired-Difference Confidence Interval for Large Sample: Use Z

SOLUTION

(a) Each subject represents a block that produces one measurement of percent body fat at the start of the semester (x) and one at the end (y). The 81 paired differences $d_i = x_i - y_i$ are summarized using a computer.

	N	Mean	StDev	SE Mean
Difference	81	3.322	2.728	0.303

That is, $\overline{d} = 3.322$ and $s_D = 2.728$. The sample size 81 is large so there is no need to assume that the population is normal. Since, from the normal table, $z_{.01} = 2.33$, the 98% confidence interval becomes

$$\left(\overline{d} - 2.33\,\frac{s_D}{\sqrt{81}},\ \overline{d} + 2.33\,\frac{s_D}{\sqrt{81}}\right)$$

$$\left(3.322 - 2.33 \times \frac{2.728}{\sqrt{81}},\ 3.322 + 2.33 \times \frac{2.728}{\sqrt{81}}\right) = (3.322 - .706,\ 3.322 + .706)$$

or (2.62, 4.03) percent. We are 98% confident that the mean reduction in body fat is 2.62 to 4.03 percent.

Large Sample Paired-Differences: Use Z-Score Paired-Difference Test

(b) Because the claim is that $\mu_D > 0$, the initial reading tends to be higher than at the end of class, we formulate:

$$H_0 : \mu_D = 0 \quad \text{versus} \quad H_1 : \mu_D > 0$$

The test statistic

$$Z = \frac{\overline{D}}{S_D / \sqrt{n}}$$

is approximately normally distributed so the rejection region is $R : Z \geq z_{.01} = 2.33$. The observed value of the test statistic

$$z = \frac{\overline{d}}{S_D / \sqrt{81}} = \frac{3.322}{2.728 / \sqrt{81}} = \frac{3.322}{.303} = 10.96$$

falls in the rejection region. Consequently H_0 is rejected in favor of H_1 at level $\alpha = .01$. We conclude that the conditioning class does reduce the mean percent body fat. The value of the test statistic is so far in the rejection region that the P-value is .0000 to at least four places. The evidence in support of H_1 is very very strong.

Paired-Difference Confidence Interval and Hypothesis Test for Small Samples: Use t

5.2 INFERENCES FROM A SMALL NUMBER OF MATCHED PAIRS

When the sample size is not large, we make the additional assumption that the distribution of the differences is normal.
In summary,

Small Samples Inferences about the Mean Difference μ_D

Assume that the differences $D_i = X_i - Y_i$ are a random sample from an $N(\mu_D, \sigma_D)$ distribution. Let

$$\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n} \quad \text{and} \quad S_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \overline{D})^2}{n-1}}$$

Then:

1. A $100(1-\alpha)\%$ confidence interval for μ_D is given by

$$\left(\overline{D} - t_{\alpha/2}\frac{S_D}{\sqrt{n}}, \quad \overline{D} + t_{\alpha/2}\frac{S_D}{\sqrt{n}}\right)$$

   where $t_{\alpha/2}$ is based on $n-1$ degrees of freedom.

2. A test of $H_0: \mu_D = \mu_{D0}$ is based on the test statistic

$$T = \frac{\overline{D} - \mu_{D0}}{S_D/\sqrt{n}} \quad \text{d.f.} = n-1$$

Small Samples Inferences about the Mean Difference δ

Assume that the differences $D_i = X_i - Y_i$ are a random sample from an $N(\delta, \sigma_D)$ distribution. Let

$$\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n} \quad \text{and} \quad S_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \overline{D})^2}{n - 1}}$$

Then:

1. A $100(1 - \alpha)\%$ confidence interval for δ is given by

$$\left(\overline{D} - t_{\alpha/2} \frac{S_D}{\sqrt{n}}, \quad \overline{D} + t_{\alpha/2} \frac{S_D}{\sqrt{n}} \right)$$

   where $t_{\alpha/2}$ is based on $n - 1$ degrees of freedom.

2. A test of $H_0 : \delta = \delta_0$ is based on the test statistic

$$T = \frac{\overline{D} - \delta_0}{S_D / \sqrt{n}} \qquad \text{d.f.} = n - 1$$

Box on Page 423

Small Samples Inferences about the Mean Difference δ

Paired-Difference Confidence Interval and Hypothesis Test for Small Samples: Use t

Example 14

Does a Pill Incidentally Reduce Blood Pressure?

A medical researcher wishes to determine if a pill has the undesirable side effect of reducing the blood pressure of the user. The study involves recording the initial blood pressures of 15 college-age women. After they use the pill regularly for six months, their blood pressures are again recorded. The researcher wishes to draw inferences about the effect of the pill on blood pressure from the observations given in Table 1.

TABLE 1 Blood-Pressure Measurements before and after Use of Pill

	Subject														
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Before (x)	70	80	72	76	76	76	72	78	82	64	74	92	74	68	84
After (y)	68	72	62	70	58	66	68	52	64	72	74	60	74	72	74
$d = x - y$	2	8	10	6	18	10	4	26	18	−8	0	32	0	−4	10

Courtesy of a family planning clinic.

(a) Calculate a 95% confidence interval for the mean reduction in blood pressure.

(b) Do the data substantiate the claim that use of the pill reduces blood pressure? Test at $\alpha = .01$.

TABLE 1 Blood-Pressure Measurements before and after Use of Pill

	Subject														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Before (x)	70	80	72	76	76	76	72	78	82	64	74	92	74	68	84
After (y)	68	72	62	70	58	66	68	52	64	72	74	60	74	72	74
$d = x - y$	2	8	10	6	18	10	4	26	18	-8	0	32	0	-4	10

Courtesy of a family planning clinic.

Table 1 (p. 423)

Blood-Pressure Measurements before and after Use of Pill

Paired-Difference Confidence Interval for Small Sample: Use t

SOLUTION

(a) Here each subject represents a block generating a pair of measurements: one before using the pill and the other after using the pill. The paired differences $d_i = x_i - y_i$ are computed in the last row of Table 1, and we calculate the summary statistics

$$\overline{d} = \frac{\sum d_i}{15} = 8.80 \quad s_D = \sqrt{\frac{\sum (d_i - \overline{d})^2}{14}} = 10.98$$

If we assume that the paired differences constitute a random sample from a normal population $N(\mu_D, \sigma_D)$, a 95% confidence interval for the mean difference μ_D is given by

$$\overline{D} \pm t_{.025} \frac{S_D}{\sqrt{15}}$$

where $t_{.025}$ is based on $\text{d.f.} = 14$. From the t table, we find $t_{.025} = 2.145$. The 95% confidence interval is then computed as

$$8.80 \pm 2.145 \times \frac{10.98}{\sqrt{15}} = 8.80 \pm 6.08 \text{ or } (2.72,\ 14.88)$$

This means that we are 95% confident the mean reduction of blood pressure is between 2.72 and 14.88.

Paired-Difference Confidence Interval for Small Sample: Use t

(b) Because the claim is that $\mu_D > 0$, we formulate

$$H_0: \mu_D = 0 \quad \text{versus} \quad H_1: \mu_D > 0$$

We employ the test statistic $T = \dfrac{\overline{D}}{S_D/\sqrt{n}}$, d.f. = 14 and set a right-sided rejection region. With d.f. = 14, we find $t_{.01} = 2.624$, so the rejection region is $R: T \geq 2.624$.

The observed value of the test statistic

$$t = \frac{\overline{d}}{S_D/\sqrt{n}} = \frac{8.80}{10.98/\sqrt{15}} = \frac{8.80}{2.84} = 3.10$$

falls in the rejection region. Consequently, H_0 is rejected in favor of H_1 at $\alpha = .01$. We conclude that a reduction in blood pressure following use of the pill is strongly supported by the data.

Paired-Difference Confidence Interval for Small Sample: Use t & Corresponding p-value

Further, as apparent from Figure 8, the observed value $t = 3.10$ has a small P-value. A software calculation gives

$$P\text{–value} = P[\,T \geq 3.10\,] = .004$$

Note: To be more convinced that the pill causes the reduction in blood pressure, it is advisable to measure the blood pressures of the same subjects once again after they have stopped using the pill for a period of time. This amounts to performing the experiment in reverse order to check the findings of the first stage.

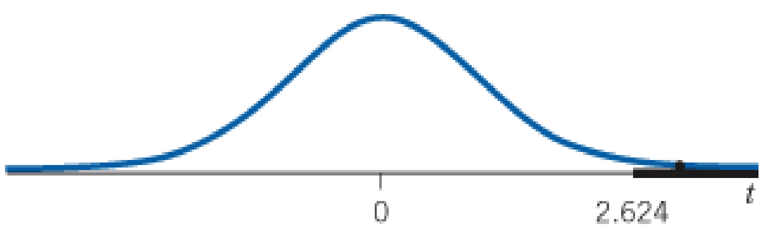


Figure 8 Rejection region $R: T \geq 2.624$ and the observed value $t = 3.10$.

Randomization with Matched Pairs: Chance Assignment - No Selection Bias

5.3 RANDOMIZATION WITH MATCHED PAIRS

Example 14 is a typical before-after situation. Data gathered to determine the effectiveness of a safety program or an exercise program would have the same structure. In such cases, there is really no way to choose how to order the experiments within a pair. The before situation must precede the after situation. If something other than the institution of the program causes performance to improve, the improvement will be incorrectly credited to the program. However, when the order of the application of treatments can be determined by the investigator, something can be done about such systematic influences. Suppose that a coin is flipped to select the treatment for the first unit in each pair. Then the other treatment is applied to the second unit. Because the coin is flipped again for each new pair, any uncontrolled variable has an equal chance of helping the performance of either treatment 1 or treatment 2. After eliminating an identified source of variation by pairing, we return to randomization in an attempt to reduce the systematic effects of any uncontrolled sources of variation.

Randomization with Pairing

After pairing, the assignment of treatments should be randomized for each pair.

Box on Page 425

Determining Whether Independent Samples and Matched Pairs Samples

6. Choosing Between Independent Samples and a Matched Pairs Sample

When planning an experiment to compare two treatments, we often have the option of either designing two independent samples or designing a sample with paired observations. Therefore, some comments about the pros and cons of these two sampling methods are in order here. Because a paired sample with n pairs of observations contains $2n$ measurements, a comparable situation would be two independent samples with n observations in each. First, note that the sample mean difference is the same whether or not the samples are paired. This is because

$$\overline{D} = \frac{1}{n}\sum (X_i - Y_i) = \overline{X} - \overline{Y}$$

Therefore, using either sampling design, the confidence intervals for the difference between treatment effects have the common form

$$(\overline{X} - \overline{Y}) \pm t_{\alpha/2} \text{ (estimated standard error)}$$

Determining Whether Independent Samples and Matched Pairs Samples

However, the estimated standard error as well as the degrees of freedom for t are different between the two situations.

	Independent Samples ($n_1 = n_2 = n$)	Paired Sample (n Pairs)
Estimated standard error	$S_{\text{pooled}}\sqrt{\frac{1}{n}+\frac{1}{n}}$	$\frac{S_D}{\sqrt{n}}$
d.f. of t	$2n-2$	$n-1$

Because the length of a confidence interval is determined by these two components, we now examine their behavior under the two competing sampling schemes.

Paired sampling results in a loss of degrees of freedom and, consequently, a larger value of $t_{\alpha/2}$. For instance, with a paired sample of $n = 10$, we have $t_{.05} = 1.833$ with d.f. $= 9$. But the t value associated with independent samples, each of size 10, is $t_{.05} = 1.734$ with d.f. $= 18$. Thus, if the estimated standard errors are equal, then a loss of degrees of freedom tends to make confidence intervals larger for paired samples. Likewise, in testing hypotheses, a loss of degrees of freedom for the t test results in a loss of power to detect real differences in the population means.

Variance Difference Smaller in Effective Pairing vs. Independent Random Samples

The merit of paired sampling emerges when we turn our attention to the other component. If experimental units are paired so that an interfering factor is held nearly constant between members of each pair, the treatment responses X and Y within each pair will be equally affected by this factor. If the prevailing condition in a pair causes the X measurement to be large, it will also cause the corresponding Y measurement to be large and vice versa. As a result, the variance of the difference $X - Y$ will be smaller in the case of an effective pairing than it will be in the case of independent random variables. The estimated standard deviation will be typically smaller as well. With an effective pairing, the reduction in the standard deviation usually more than compensates for the loss of degrees of freedom.

Paired-Sampling Effective in Reducing Variability Among Study Subjects. But if Study Subjects Already Alike May Not Reduce Variance. Result is Loss of Degrees of Freedom w/Paired Difference

In Example 14, concerning the effect of a pill in reducing blood pressure, we note that a number of important factors (age, weight, height, general health, etc.) affect a person's blood pressure. By measuring the blood pressure of the same person before and after use of the pill, these influencing factors can be held nearly constant for each pair of measurements. On the other hand, independent samples of one group of persons using the pill and a separate control group of persons not using the pill are apt to produce a greater variability in blood-pressure measurements if all the persons selected are not similar in age, weight, height, and general health.

In summary, paired sampling is preferable to independent sampling when an appreciable reduction in variability can be anticipated by means of pairing. When the experimental units are already alike or their dissimilarities cannot be linked to identifiable factors, an arbitrary pairing may fail to achieve a reduction in variance. The loss of degrees of freedom will then make a paired comparison less precise.

Review: Simultaneous Confidence Intervals for Mean Differences in ANOVA

Simultaneous Confidence Intervals for Mean Differences of Groups in ANOVA

3. Simultaneous Confidence Intervals

The ANOVA F test is only the initial step in our analysis. It determines if significant differences exist among the treatment means. Our goal should be more than to merely conclude that treatment differences are indicated by the data. Rather, we must detect likenesses and differences among the treatments. The problem of estimating differences in treatment means is of even greater importance than the overall F test.

Referring to the comparison of k treatments using the data structure given in Table 1, let us examine how a confidence interval can be established for $\mu_1 - \mu_2$, the mean difference between treatment 1 and treatment 2. The statistic

$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\text{SSE}}{n-k}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has a t distribution with $\text{d.f.} = n - k$, and this can be employed to construct a confidence interval for $\mu_1 - \mu_2$.

Confidence Interval for Single Difference Between Two Means

More generally:

Confidence Interval for a Single Difference

A $100\,(1-\alpha)\,\%$ confidence interval for $\mu_i - \mu_{i'}$, the difference of means for treatment i and treatment i' is given by

$$\left(\bar{Y}_i - \bar{Y}_{i'}\right) \pm t_{\alpha/2}\, S\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}$$

where

$$S = \sqrt{\mathrm{MSE}} = \sqrt{\frac{\mathrm{SSE}}{n-k}}$$

and $t_{\alpha/2}$ is the upper $\alpha/2$ point of t with d.f. $= n-k$.

Even though the confidence interval is for the difference of two particular means, the mean square error is computed using the data from all k treatments. This is because we assume that the variance is the same for all treatments.

Multiple Confidence Intervals for Difference Between Several Means

If the F test first shows a significant difference in means, then some statisticians feel that it is reasonable to compare means pairwise according to the preceding intervals. However, many statisticians prefer a more conservative procedure based on the following reasoning.

Without the provision that the F test is significant, the preceding method provides **individual** confidence intervals for pairwise differences. However, with $k = 4$ treatments, there are $\binom{4}{2} = 6$ pairwise differences $\mu_i - \mu_{i'}$, and this procedure applied to all pairs yields six confidence statements, each having a $100\,(1-\alpha)\,\%$ level of confidence. It is difficult to determine what level of confidence will be achieved for claiming that **all** six of these statements are correct. To overcome this dilemma, procedures have been developed for several confidence intervals to be constructed in such a manner that the joint probability that all the statements are true is guaranteed not to fall below a predetermined level. Such intervals are called **multiple confidence intervals** or **simultaneous confidence intervals.** Numerous methods proposed in the statistical literature have achieved varying degrees of success. We present one that can be used simply and conveniently in general applications.

Multiple-t Confidence Intervals

The procedure, called the **multiple-t confidence intervals,** consists of setting confidence intervals for the differences $\mu_i - \mu_{i'}$ in much the same way we just did for the individual differences, except that a different percentage point is read from the t table.

Operationally, the construction of these confidence intervals does not require any new concepts or calculations, but it usually involves some nonstandard percentage point of t. For example, with $k = 3$ and $1 - \alpha = .95$, if we want to set simultaneous intervals for all $m = \binom{k}{2} = 3$ pairwise differences, we require that the upper $\alpha / (2m) = .05/6 = .00833$ point of a t distribution.

Multiple-t Confidence Intervals

A set of $100(1-\alpha)\%$ simultaneous confidence intervals for m number of pairwise differences $\mu_i - \mu_{i'}$ is given by

$$(\bar{Y}_i - \bar{Y}_{i'}) \pm t_{\alpha/2m}\, S\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}$$

where $S = \sqrt{\text{MSE}}$, $m =$ the number of confidence statements, and $t_{\alpha/2m} =$ the upper $\alpha/(2m)$ point of t with d.f. $= n - k$.

Prior to sampling, the probability of all the m statements being correct is at least $1 - \alpha$.

Multiple-t Confidence Intervals

A set of $100(1 - \alpha)\%$ simultaneous confidence intervals for m number of pairwise differences $\mu_i - \mu_{i'}$ is given by

$$(\overline{Y}_i - \overline{Y}_{i'}) \pm t_{\alpha/2m} S \sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}$$

where $S = \sqrt{\text{MSE}}$, $m =$ the number of confidence statements, and $t_{\alpha/2m} =$ the upper $\alpha/(2m)$ point of t with d.f. $= n - k$.

Prior to sampling, the probability of all the m statements being correct is at least $1 - \alpha$.

Box on Page 574

Multiple Confidence Intervals

Example: Calculating Multiple-t Confidence Intervals to Determine Which Means Differ

Example 4

Calculating Multiple-*t* Confidence Intervals to Reveal Which means Differ

An experiment is conducted to determine the soil moisture deficit resulting from varying amounts of residual timber left after cutting trees in a forest. The three treatments are treatment 1: no timber left; treatment 2: 2000 bd ft left; treatment 3: 8000 bd ft left. (Board feet is a particular unit of measurement of timber volume.) The measurements of moisture deficit are given in Table 8. Perform the ANOVA test and construct 95% multiple-*t* confidence intervals for the treatment differences.

TABLE 8 Moisture Deficit in Soil

Treatment	Observations	Total	Mean
1	1.52 1.38 1.29 1.48 1.63	$T_1 = 7.30$	$\bar{y}_1 = 1.460$
2	1.63 1.82 1.35 1.03 2.30 1.45	$T_2 = 9.58$	$\bar{y}_2 = 1.597$
3	2.56 3.32 2.76 2.63 2.12 2.78	$T_3 = 16.17$	$\bar{y}_3 = 2.695$
		Grand total	Grand mean
		$T = 33.05$	$\bar{y} = 1.944$

TABLE 8 Moisture Deficit in Soil

Treatment	Observations						Total	Mean
1	1.52	1.38	1.29	1.48	1.63		$T_1 = 7.30$	$\bar{y}_1 = 1.460$
2	1.63	1.82	1.35	1.03	2.30	1.45	$T_2 = 9.58$	$\bar{y}_2 = 1.597$
3	2.56	3.32	2.76	2.63	2.12	2.78	$T_3 = 16.17$	$\bar{y}_3 = 2.695$
							Grand total	Grand mean
							$T = 33.05$	$\bar{y} = 1.944$

Table 8 (p. 575)

Moisture Deficit in Soil

Example: Calculating Multiple-t Confidence Intervals to Determine Which Means Differ

SOLUTION

Our analysis employs convenient alternative forms of the expressions for sums of squares involving totals. The total number of observations $n = 5 + 6 + 6 = 17$.

$$\begin{aligned}
\text{Total SS} &= \sum_{i=1}^{3}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}\right)^2 = \sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{n} \\
&= 71.3047 - 64.2531 = 7.0516 \\
\text{Treatment SS} &= \sum_{i=1}^{3} n_i \left(\bar{y}_i - \bar{y}\right)^2 = \sum_{i=1}^{3} \frac{T_i^2}{n_i} - \frac{T^2}{n} \\
&= 69.5322 - 64.2531 = 5.2791 \\
\text{Error SS} &= \text{Total SS} - \text{Treatment SS} = 1.7725
\end{aligned}$$

The ANOVA table appears in Table 9.

TABLE 9 ANOVA Table for Comparison of Moisture Deficit

Source	Sum of Squares	d.f.	Mean Square	*F*-ratio
Treatment	5.2791	2	2.640	20.8
Error	1.7725	14	.127	
Total	7.0516	16		

Example: Calculating Multiple-t Confidence Intervals – Use Combinations Rule to Determine How Many Pair Differences

TABLE 9 ANOVA Table for Comparison of Moisture Deficit

Source	Sum of Squares	d.f.	Mean Square	*F*-ratio
Treatment	5.2791	2	2.640	20.8
Error	1.7725	14	.127	
Total	7.0516	16		

Because the observed value of F is larger than the tabulated value $F_{.05}(2, 14) = 3.74$, the null hypothesis of no difference in the treatment effects is rejected at $\alpha = .05$. The P-value is less than .0001 which provides very strong evidence against the null hypothesis.

In constructing a set of 95% multiple-t confidence intervals for pairwise differences, note that there are $\binom{3}{2} = 3$ pairs, so

$$\frac{\alpha}{2m} = \frac{.05}{(2 \times 3)} = .00833$$

TABLE 9 ANOVA Table for Comparison of Moisture Deficit

Source	Sum of Squares	d.f.	Mean Square	F-ratio
Treatment	5.2791	2	2.640	20.8
Error	1.7725	14	.127	
Total	7.0516	16		

Table 9 (p. 575)

ANOVA Table for Comparison of Moisture Deficit

Use t value to Calculate Confidence Interval Between Pairs of Means

In constructing a set of 95% multiple-t confidence intervals for pairwise differences, note that there are $\binom{3}{2} = 3$ pairs, so

$$\frac{\alpha}{2m} = \frac{.05}{(2 \times 3)} = .00833$$

From Appendix B, Table 6, the upper .00833 point of t with d.f. = 14 is 2.718. The simultaneous confidence intervals are calculated as follows:

$$\mu_2 - \mu_1: \quad (1.597 - 1.460) \pm 2.718 \times .356 \times \sqrt{\frac{1}{6} + \frac{1}{5}}$$
$$= (-.45, .72)$$

$$\mu_3 - \mu_2: \quad (2.695 - 1.597) \pm 2.718 \times .356 \times \sqrt{\frac{1}{6} + \frac{1}{6}}$$
$$= (.54, 1.66)$$

$$\mu_3 - \mu_1: \quad (2.695 - 1.460) \pm 2.718 \times .356 \times \sqrt{\frac{1}{6} + \frac{1}{5}}$$
$$= (.65, 1.82)$$

These confidence intervals indicate that treatments 1 and 2 do not differ appreciably, but the mean for treatment 3 is considerably higher than the means for treatments 1 and 2.

Comparing Two Samples: Variance or Standard Deviations Are Equal

Comparing Two Samples: Variances σ^2 or Standard Deviation σ Are Equal

3. Inferences About the Difference of Means—Independent Small Samples From Normal Populations

Not surprisingly, more distributional structure must be imposed before inference procedures can be developed for small samples. Here, we require that both population distributions be normal. A restriction to normal populations is not new and was already required for small sample inferences about a single mean. We treat the equal variance case first. It is the simplest.

3.1 INFERENCES: NORMAL POPULATIONS WITH EQUAL VARIANCES

The additional assumptions specify that the normal populations are normal and they also require the two standard deviations to be equal.

Additional Assumptions When the Sample Sizes Are Small

1. Both populations are normal.
2. The population standard deviations σ_1 and σ_2 are equal.

Additional Assumptions When the Sample Sizes Are Small

1. Both populations are normal.
2. The population standard deviations σ_1 and σ_2 are equal.

Box on Page 405

Small Sample Assumptions: Pool Variance and Then Divide by n_1 and n_2

The assumption requiring equal variability is somewhat artificial but we reserve comment until later. Letting σ denote the common standard deviation, we summarize.

Small Samples Assumptions

1. $X_1, X_2, \ldots, X_{n_1}$ is a random sample from $N(\mu_1, \sigma)$.
2. $Y_1, Y_2, \ldots, Y_{n_2}$ is a random sample from $N(\mu_2, \sigma)$.

   (*Note*: σ is the same for both distributions.)
3. $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are independent.

Estimation

Again, $\overline{X}-\overline{Y}$ is our choice for a statistic.

$$\text{Mean of } (\overline{X}-\overline{Y}) = E(\overline{X}-\overline{Y}) = \mu_1 - \mu_2$$

$$\text{Var}(\overline{X}-\overline{Y}) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

Pooled Variance or S^2: Numerator and Denominator Include Degrees of Freedom for Both Samples n_1 and n_2

The common variance σ^2 can be estimated by combining information provided by both samples. Specifically, the sum $\sum_{i=1}^{n_1} (X_i - \overline{X})^2$ incorporates $n_1 - 1$ pieces of information about σ^2, in view of the constraint that the deviations $X_i - \overline{X}$ sum to zero. Independently of this, $\sum_{i=1}^{n_2} (Y_i - \overline{Y})^2$ contains $n_2 - 1$ pieces of information about σ^2. These two quantities can then be combined,

$$\sum (X_i - \overline{X})^2 + \sum (Y_i - \overline{Y})^2$$

to obtain a pooled estimate of the common σ^2. The proper divisor is the sum of the component degrees of freedom, or $(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$.

Pooled Estimator of the Common σ^2

$$S^2_{\text{pooled}} = \frac{\sum_{i=1}^{n_1} (X_i - \overline{X})^2 + \sum_{i=1}^{n_2} (Y_i - \overline{Y})^2}{n_1 + n_2 - 2}$$

$$= \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}$$

Small Samples Assumptions

1. $X_1, X_2, \ldots, X_{n_1}$ is a random sample from $N(\mu_1, \sigma)$.
2. $Y_1, Y_2, \ldots, Y_{n_2}$ is a random sample from $N(\mu_2, \sigma)$.
   (*Note:* σ is the same for both distributions.)
3. $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are independent.

Pooled Estimator of the Common σ^2

$$S^2_{\text{pooled}} = \frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{i=1}^{n_2}(Y_i - \overline{Y})^2}{n_1 + n_2 - 2}$$

$$= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Boxes on Pages 405, 406

Small Samples Assumptions; Pooled Estimator of the Common σ^2

Small Samples Confidence Interval Calculation of Two Samples: Pooled Variance

Example 9

Calculating a Small Samples Confidence Interval

Beginning male and female accounting students were given a test and, on the basis of their answers, were assigned a computer anxiety score (CARS). Using the data given in Table D.4 of the Data Bank, obtain a 95% confidence interval for the difference in mean computer anxiety score between beginning male and female accounting students.

SOLUTION

The dot diagrams of these data, plotted in Figure 6, give the appearance of approximately equal amounts of variation.

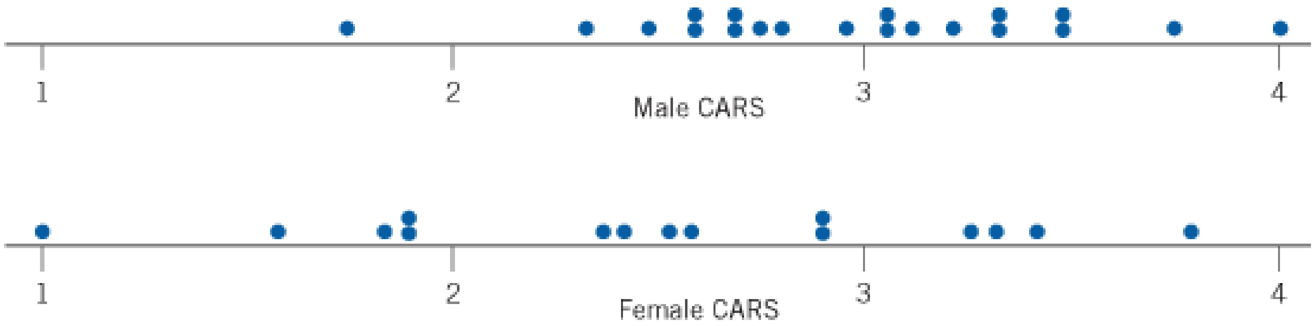


Figure 6 Dot diagrams of the computer anxiety data in Example 9.

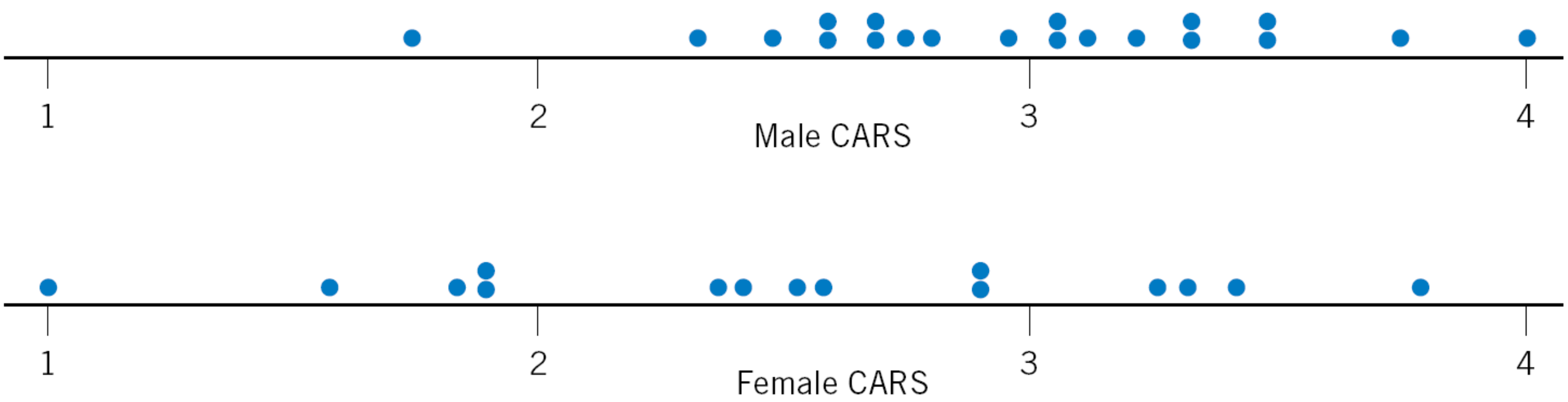


Figure 6 (p. 408)

Dot diagrams of the computer anxiety data in Example 9.

Confidence Interval for Two Samples: Calculate Mean and Standard Deviation Respectively for Each

We assume that the CARS data for both females and males are random samples from normal populations with means μ_1 and μ_2, respectively, and a common standard deviation σ . Computations from the data provide the summary statistics:

Female CARS

$$n_1 = 15 \quad \overline{x} = 2.514 \quad s_1 = .773$$

Male CARS

$$n_2 = 20 \quad \overline{y} = 2.963 \quad s_2 = .525$$

We calculate

$$S_{\text{pooled}} = \sqrt{\frac{14\,(.773)^2 + 19\,(.525)^2}{15 + 20 - 2}} = .642$$

Confidence Interval for Two Samples: Use t critical Value Multiplied by Pooled Standard Deviation of 2 Samples

With a 95% confidence interval $\alpha/2 = .025$ and consulting the t table, we find (interpolating) that $t_{.025} = 2.035$ for $\text{d.f.} = n_1 + n_2 - 2 = 33$. Thus a 95% confidence interval for $\mu_1 - \mu_2$ becomes

$$\bar{x} - \bar{y} \pm t_{0.25}\, s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 2.514 - 2.963 \pm 2.035 \times .642 \sqrt{\frac{1}{15} + \frac{1}{20}}$$

$$= -.449 \pm .446 \quad \text{or} \quad (-.895, -.003)$$

We can be 95% confident that the mean computer anxiety score for female beginning accounting students can be .003 to .895 units lower than the mean score for males.

This interval is quite wide. Certainly the values near $-.003$ represent technically insignificant differences.

Confidence Interval for $\mu_1 - \mu_2$ Small Samples

A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\bar{X} - \bar{Y} \pm t_{\alpha/2} \, S_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where

$$S^2_{\text{pooled}} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

and $t_{\alpha/2}$ is the upper $\alpha/2$ point of the t distribution with d.f. $= n_1 + n_2 - 2$.

Box on Page 407

Confidence Interval for $\mu_1 - \mu_2$.

Hypothesis Tests for Two Samples: Subtract the Means and Formulate Ho and H_1 or H_a

Tests of Hypotheses

Tests of hypotheses concerning the difference in means are based on a statistic having student's t distribution.

Testing H_0: $\mu_1 - \mu_2 = \delta_0$ with Small Samples and $\sigma_1 = \sigma_2$

Test statistic:

$$T = \frac{(\overline{X} - \overline{Y}) - \delta_0}{S_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{d.f.} = n_1 + n_2 - 2$$

Alternative Hypothesis	Level α Rejection Region
$H_1: \mu_1 - \mu_2 > \delta_0$	$R: T \geq t_\alpha$
$H_1: \mu_1 - \mu_2 < \delta_0$	$R: T \leq -t_\alpha$
$H_1: \mu_1 - \mu_2 \neq \delta_0$	$R: \lvert T \rvert \geq t_{\alpha/2}$

Testing H_0: $\mu_1 - \mu_2 = \delta_0$ with Small Samples and $\sigma_1 = \sigma_2$

Test statistic:

$$T = \frac{(\overline{X} - \overline{Y}) - \delta_0}{S_{\text{pooled}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad \text{d.f.} = n_1 + n_2 - 2$$

Alternative Hypothesis	Level α Rejection Region
$H_1: \mu_1 - \mu_2 > \delta_0$	$R: T \geq t_\alpha$
$H_1: \mu_1 - \mu_2 < \delta_0$	$R: T \leq -t_\alpha$
$H_1: \mu_1 - \mu_2 \neq \delta_0$	$R: \lvert T \rvert \geq t_{\alpha/2}$

Box on Page 409

Testing H_0: $\mu_1 - \mu_2 = \delta_0$ with small samples

Hypothesis Testing 2 Samples Equality of Means: Use t-Test With Pooled Square Root of Variance i.e. Standard Deviation

Example 10

Testing the Equality of Mean Computer Anxiety Scores

Refer to the computer anxiety scores (CARS) described in Example 9 and the summary statistics

Female CARS

$$n_1 = 15 \quad \bar{x} = 2.514 \quad s_1 = .773$$

Male CARS

$$n_2 = 20 \quad \bar{y} = 2.963 \quad s_2 = .525$$

Do these data strongly indicate that the mean score for females is lower than that for males? Test at level $\alpha = .05$.

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_{\text{pooled}} \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

has Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom.

Box on Page 407

Hypothesis Testing 2 Samples t-Test With Standard Deviation: To Find t Critical Value Must Pool Degrees of Freedom

SOLUTION

We are seeking strong evidence in support of the hypothesis that the mean computer anxiety score for females (μ_1) is less than the mean score for males. Therefore the alternative hypothesis should be taken as $H_1: \mu_1 < \mu_2$ or $H_1: \mu_1 - \mu_2 < 0$, and our problem can be stated as testing

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{versus} \quad H_1: \mu_1 - \mu_2 < 0$$

We employ the test statistic

$$T = \frac{\overline{X} - \overline{Y}}{S_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{d.f.} = n_1 + n_2 - 2$$

and set the left-sided rejection region $R: T \le -t_{.05}$. For d.f. $= n_1 + n_2 - 2 = 33$, we approximate the tabled value as $t_{.05} = 1.692$. The rejection region is $R: T \le -1.692$ as illustrated in Figure 7.

Right-Sided Hypothesis Test of Mean Anxiety Score Females Less Than Males: Compare Calculated t to t Critical Value

and set the left-sided rejection region $R: T \leq -t_{.05}$. For d.f. $= n_1 + n_2 - 2 = 33$, we approximate the tabled value as $t_{.05} = 1.692$. The rejection region is $R: T \leq -1.692$ as illustrated in Figure 7.

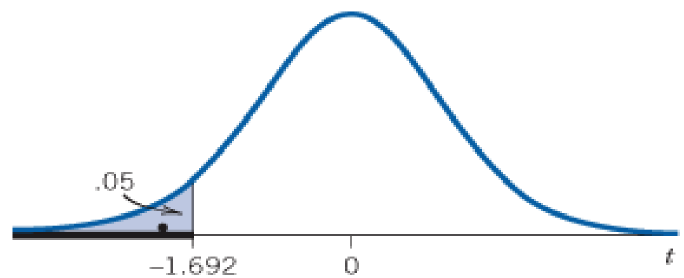


Figure 7 Rejection region for Example 10.

With $S_{\text{pooled}} = .642$ already calculated in Example 9, the observed value of the test statistic T is

$$t = \frac{2.514 - 2.963}{.642\sqrt{\frac{1}{15} + \frac{1}{20}}} = \frac{-.449}{.2193} = -2.05$$

This value lies in the rejection region R. Consequently, at the .05 level of significance, we reject the null hypothesis in favor of the alternative hypothesis that males have a higher mean computer anxiety score.

A computer calculation gives a P-value of about .025 so the evidence of H_1 is moderately strong.

Confidence Interval for 2 Samples With Unequal Variances: Using Pooled σ

Example 11

A Confidence Interval Shows Mothers of Males Have Milk with Higher Energy Content

Researchers, interested in the energy content of mothers' milk, conducted a study with 25 well-nourished, healthy mothers exclusively breast feeding their 2- to 5-month old babies. All of the participants lived in Massachusetts.[3] C. E. Powe, C. D. Knott, and N. Conklin-Brittain, "*Infant sex predicts breast milk energy content*," *American Journal of Human Biology* 22 (2010), pp. 50-54. Thirteen of the mothers had sons and twelve had daughters. The summary statistics

Energy (kcal/100 ml)

	Males	**Females**
Number	13	12
Mean	75.56	60.81
Standard deviation	19.37	15.64

are obtained in the study.

Does the gender of baby effect the amount of energy available in the mothers' milk? If so, what can you say about the mean difference?

Base your answers on a 95% confidence interval for the difference of mean energy density between the mothers' milk of male babies and and the mothers' milk of female babies.

Confidence Interval for 2 Samples With Unequal Variances: Using Pooled σ

SOLUTION

Since the two standard deviations are not very different, we use the pooled estimate of variance

$$s_{\text{pooled}} = \sqrt{\frac{13\,(19.37)^2 + 12\,(15.64)^2}{13 + 12 - 2}} = 17.684$$

For a 95% confidence interval, $\alpha / 2 = .025$, and we find $t_{.025} = 2.069$ for d.f. $= 23$. The resulting 95% confidence interval is

$$\begin{aligned} \bar{x} - \bar{y} \;\; &\pm \;\; t_{.025}\, s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ &= 75.56 - 60.81 \pm 2.069 \times 17.684 \sqrt{\frac{1}{13} + \frac{1}{12}} \\ &= 14.75 \pm 14.647 \end{aligned}$$

or $.10 < \mu_1 - \mu_2 < 29.40$

Confidence Interval for 2 Samples With Unequal Standard Deviation: Using Pooled σ

The 95% confidence interval includes only positive values for $\mu_1 - \mu_2$. With 95% confidence, we conclude that the mean energy density is larger when the baby is male than when it is female. However, the mean difference could be almost as small as .10 kcal/100ml which is likely not scientifically significant. It also could be almost as large as 29.4 kcal/100ml which is an important difference. A larger study is required to more accurately assess the mean difference.

This study involved well nourished healthy mothers in Massachusetts but they were not a random sample of new mothers. Generalizations are difficult, especially extrapolations to under-nourished mothers.

Pooling Variance vs. Not Pooling Standard Deviation When Constructing Confidence Intervals

Deciding Whether or not to Pool

Our preceding discussion of large and small sample inferences raises a few questions:

For small sample inference, why do we assume the population standard deviations to be equal when no such assumption was needed in the large samples case?

When should we be wary about this assumption, and what procedure should we use when the assumption is n

reasonable?

Learning statistics would be a step simpler if the ratio

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

had a t distribution for small samples from normal populations. Unfortunately, statistical theory proves it otherwise. The distribution of this ratio is *not* a t and, worse yet, it depends on the unknown quantity σ_1/σ_2.

Pooled Standard Deviation Rule of Thumb: If S_1/S_2 smaller than ½ or Greater than 2, Do NOT Pool

The assumption $\sigma_1 = \sigma_2$ and the change of the denominator to $S_{\text{pooled}}\sqrt{1/n_1 + 1/n_2}$ allow the t-based inferences to be valid. However, the $\sigma_1 = \sigma_2$ restriction and accompanying pooling are not needed in large samples where a normal approximation holds.

With regard to the second question, the relative magnitude of the two sample standard deviations s_1 and s_2 should be the prime consideration. The assumption $\sigma_1 = \sigma_2$ is reasonable if s_1/s_2 is not very much different from 1. As a working rule, the range of values $\frac{1}{2} \le s_1 \big/ s_2 \le 2$ may be taken as reasonable cases for making the assumption $\sigma_1 = \sigma_2$ and hence for pooling. If s_1/s_2 is seen to be smaller than $\frac{1}{2}$ or larger than 2, the assumption $\sigma_1 = \sigma_2$ would be suspect. In that case, some approximate methods of inference about $\mu_1 - \mu_2$ are available.

Review: 2 Samples With Unequal Variances or Standard Deviations

Two Samples with Unequal Variances or Standard Deviations i.e. σ

3.2 INFERENCES: NORMAL POPULATIONS WITH UNEQUAL VARIANCES

When the sample sizes are not large and the observed standard deviations quite different, we can still proceed to test hypotheses and obtain confidence intervals provided that the populations are normal. The procedure we introduce is somewhat complex but is widely used in most statistical software programs. Consequently, using software is the preferred approach.

Inference procedures are developed based on a statistic that is the same as the large sample statistic for comparing two means. However, we use the notation T^*, where

$$T^* = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Because the sample sizes are small, we chose to approximate the distribution of T^* by a t distribution. But, its degrees of freedom are estimated using the sample standard deviations. As a consequence, the distribution of T^* is only approximately a t distribution.

Estimating Degrees of Freedom for Two Samples with Unequal Standard Deviation

This approach of approximating the distribution of T^* by a t, is sometimes called the **Satterthwaite correction** after the person who derived the approximation.

Normal Populations with $\sigma_1 \neq \sigma_2$

Let $\delta = \mu_1 - \mu_2$. When the sample sizes n_1 and n_2 are not large and $\sigma_1 \neq \sigma_2$

$$T^* = \frac{\overline{X} - \overline{Y} - \delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

is approximately distributed as a t with estimated degrees of freedom.

The estimated degrees of freedom depend on the observed values of the sample variances s_1^2 and s_2^2.

$$\text{estimated degrees of freedom} = \nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

The estimated degrees of freedom ν are often rounded down to an integer so a t table can be consulted.

Normal Populations with $\sigma_1 \neq \sigma_2$

Let $\delta = \mu_1 - \mu_2$. When the sample sizes n_1 and n_2 are not large and $\sigma_1 \neq \sigma_2$

$$T^* = \frac{\overline{X} - \overline{Y} - \delta}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

is approximately distributed as a t with estimated degrees of freedom.

The estimated degrees of freedom depend on the observed values of the sample variances s_1^2 and s_2^2.

estimated degrees of freedom =

$$v = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2} \right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{s_1^2}{n_1} \right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{s_2^2}{n_2} \right)^2}$$

Box on Page 413

Small Sample Inferences for μ_1 - μ_2 When the Populations Are Normal But σ_1 or σ_2 Are Not Assumed to Be Equal

Hypothesis Testing and Confidence Interval For Two Sample Means Using Pooled Standard Deviation for Unequal Variances

Example 12

Testing Equality and Confidence Intervals Concerning Mean of Green Gas

One process of making green gasoline, not just a gasoline additive, takes biomass in the form of sucrose and converts it into gasoline using catalytic reactions. This research is still at the pilot plant stage. At one step in a pilot plant process, the product volume (liters) consists of carbon chains of length 3. Nine runs were made with each of two catalysts and the product volumes measured.

catalyst 1	1.86	2.05	2.06	1.88	1.75	1.64	1.86	1.75	2.13
catalyst 2	.32	1.32	.93	.84	.55	.84	.37	.52	.34

The sample sizes $n_1 = n_2 = 9$ and the summary statistics are

$$\overline{x} = 1.887, \quad s_1^2 = .0269 \quad \overline{y} = .670 \quad s_2^2 = .1133$$

(a) Is the mean yield with catalyst 1 more than .80 liters higher than the yield with catalyst 2? Test with $\alpha = 0.05$.

(b) Find a 98% confidence interval for the difference of means.

Hypothesis Testing For Two Sample Means Using Pooled Standard Deviation Formula for Unequal Variances

SOLUTION

(a) The test concerns the $\delta = \mu_1 - \mu_2$ and we wish to show that δ is greater than .80. Therefore we formulate the testing problem as

$$H_0: \mu_1 - \mu_2 = .80 \quad \text{versus} \quad H_1: \mu_1 - \mu_2 > .80$$

The sample sizes $n_1 = n_2 = 9$ are small and there are no outliers or obvious departures from normality

However, $s_2^2 = .1133$ is more than four times $s_1^2 = .0269$. We should not pool.

To apply the approximate t, we first estimate the degrees of freedom

$$\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{S_2^2}{n_2}\right)^2} = \frac{\left(\frac{.0269}{9} + \frac{.1133}{9}\right)^2}{\frac{1}{9-1}\left(\frac{.0269}{9}\right)^2 + \frac{1}{9-1}\left(\frac{.1133}{9}\right)^2} = 11.60$$

Confidence Interval For Two Sample Means Using Pooled Standard Deviation Formula for Unequal Variances

To use the t table, we round down to 11 and then obtain $t_{0.025} = 2.201$ for 11 degrees of freedom.

The rejection region becomes

$$T^* = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \geq 2.201$$

The calculated value of the test statistic is

$$t^* = \frac{1.887 - .670 - .80}{\sqrt{\frac{.0269}{9} + \frac{.1133}{9}}} = \frac{.417}{.1248} = 3.34$$

which falls in the rejection region. So, we reject the null hypothesis at $\alpha = .05$ and conclude that the mean product volume from catalyst 1 is more than .80 liters higher.

(b) Since $t_{.01} = 2.718$, the approximate 98% confidence interval is

$$\overline{x} - \overline{y} \pm 2.718 \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.887 - .670 \pm 2.718\ (.1248) \text{ or } (.88,\ 1.56)$$