CHAPTER 12

General Linear Hypothesis

12.1 INTRODUCTION

This chapter deals with the general linear hypothesis. In a wide variety of problems the experimenter is interested in making inferences about a vector parameter. For example, he may wish to estimate the mean of a multivariate normal or to test some hypotheses concerning the mean vector. The problem of estimation can be solved, for example, by resorting to the method of maximum likelihood estimation, discussed in Section 8.7. In this chapter we restrict ourselves to linear model problems and concern ourselves mainly with problems of hypothesis testing.

In Section 12.2 we formally describe the general model and derive a test in complete generality. In the next four sections we demonstrate the power of this test by solving four important testing problems. We need a considerable amount of linear algebra in Section 12.2.

12.2 GENERAL LINEAR HYPOTHESIS

A wide variety of problems of hypothesis testing can be treated under a general setup. In this section we state the general problem, and derive the test statistic and its distribution. Consider the following examples.

Example 1. Let $Y_1, Y_2, \ldots, Y_k$ be independent RVs with $EY_i = \mu_i$, $i = 1, 2, \ldots, k$, and common variance σ^2. Also, n_i observations are taken on Y_i, $i = 1, 2, \ldots, k$, and $\sum_{i=1}^{k} n_i = n$. It is required to test $H_0\colon \mu_i = \mu_2 = \cdots = \mu_k$. The case $k = 2$ has already been treated in Section 10.4. Problems of this nature arise quite naturally, for example, in agricultural experiments where one is interested in comparing the average yield when k fertilizers are available.

Example 2. An experimenter observes the velocity of a particle moving along a line. He takes observations at given times $t_1, t_2, \ldots, t_n$. Let β_1 be the initial velocity of the particle and β_2 be the acceleration; then the velocity at time t is given by $y = \beta_1 + \beta_2 t + \varepsilon$, where ε is an RV that is nonobservable (e.g., an error in measurement).

In practice, the experimenter does not know β_1 and β_2 and has to use the random observations $Y_1, Y_2, \ldots, Y_n$ made at times $t_1, t_2, \ldots, t_n$, respectively, to obtain some information about the unknown parameters β_1, β_2.

A similar example is the case when the relation between y and t is governed by

$$y = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon,$$

where t is a mathematical variable, $\beta_0, \beta_1, \beta_2$ are unknown parameters, and ε is a nonobservable RV. The experimenter takes observations $Y_1, Y_2, \ldots, Y_n$ at predetermined values $t_1, t_2, \ldots, t_n$, respectively, and is interested in testing the hypothesis that the relation is in fact linear, that is, $\beta_2 = 0$.

Examples of the type discussed above and their much more complicated variants can all be treated under a general setup. To fix ideas, let us first make the following definition.

Definition 1. Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$ be a random column vector and $\mathbf{X}$ be an $n \times k$ matrix, $k < n$, of known constants x_{ij}, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, k$. We say that the distribution of $\mathbf{Y}$ satisfies a *linear model* if

$$E\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ is a vector of unknown (scalar) parameters $\beta_1, \beta_2, \ldots, \beta_k$. It is convenient to write

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is a vector of nonobservable RVs with $E\varepsilon_j = 0$, $j = 1, 2, \ldots, n$. Relation (2) is known as a linear model. Then the general linear hypothesis concerns $\boldsymbol{\beta}$, namely, that $\boldsymbol{\beta}$ satisfies H_0: $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{H}$ is a known $r \times k$ matrix with $r \leq k$.

In what follows we assume that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent, normal RVs with common variance σ^2 and $E\varepsilon_j = 0$, $j = 1, 2, \ldots, n$. In view of (2), it follows that $Y_1, Y_2, \ldots, Y_n$ are independent normal RVs with

$$EY_i = \sum_{j=1}^{k} x_{ij}\beta_j \quad \text{and} \quad \operatorname{var}(Y_i) = \sigma^2, \qquad i = 1, 2, \ldots, n. \tag{3}$$

We assume that $\mathbf{H}$ is a matrix of full rank r, $r \leq k$, and $\mathbf{X}$ is a matrix of full rank $k < n$. Some remarks are in order.

Remark 1. Clearly, $\mathbf{Y}$ satisfies a linear model if the vector of means $E\mathbf{Y} = (EY_1, EY_2, \ldots, EY_n)'$ lies in a k-dimensional subspace generated by the linearly independent column vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ of the matrix $\mathbf{X}$. Indeed, (1) states that

$E\mathbf{Y}$ is a linear combination of the known vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$. The general linear hypothesis $H_0: \mathbf{H}\boldsymbol{\beta} = \mathbf{0}$ states that the parameters $\beta_1, \beta_2, \ldots, \beta_k$ satisfy r independent homogeneous linear restrictions. It follows that under H_0, $E\mathbf{Y}$ lies in a $(k-r)$-dimensional subspace of the k-space generated by $\mathbf{x}_1, \ldots, \mathbf{x}_k$.

Remark 2. The assumption of normality, which is conventional, is made to compute the likelihood ratio test statistic of H_0 and its distribution. If the problem is to estimate $\boldsymbol{\beta}$, no such assumption is needed. One can use the *principle of least squares* and estimate $\boldsymbol{\beta}$ by minimizing the sum of squares,

$$\sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \tag{4}$$

The minimizing value $\hat{\boldsymbol{\beta}}(\mathbf{y})$ is known as a *least squares estimate of* $\boldsymbol{\beta}$. This is not a difficult problem and we do not discuss it here in any detail but mention only that any solution of the *normal equations*

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y} \tag{5}$$

is a least squares estimator. If the rank of $\mathbf{X}$ is $k(<n)$, then $\mathbf{X}'\mathbf{X}$, which has the same rank as $\mathbf{X}$, is a nonsingular matrix that can be inverted to give a unique least squares estimator

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \tag{6}$$

If the rank of $\mathbf{X}$ is $< k$, then $\mathbf{X}'\mathbf{X}$ is singular and the normal equations do not have a unique solution. One can show, for example, that $\hat{\boldsymbol{\beta}}$ is unbiased for $\boldsymbol{\beta}$, and if the Y_i's are uncorrelated with common variance σ^2, the variance–covariance matrix of the $\hat{\beta}_i$'s is given by

$$E\left\{\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right\} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \tag{7}$$

Remark 3. One can similarly compute the *restricted least squares estimator* of $\boldsymbol{\beta}$ by the usual method of Lagrange multipliers. For example, under $H_0: \mathbf{H}\boldsymbol{\beta} = \mathbf{0}$, one simply minimizes $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ subject to $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$ to get the restricted least squares estimator $\hat{\hat{\boldsymbol{\beta}}}$. The important point is that if $\boldsymbol{\varepsilon}$ is assumed to be a multivariate normal RV with mean vector $\mathbf{0}$ and dispersion matrix $\sigma^2\mathbf{I}_n$, the MLE of $\boldsymbol{\beta}$ is the same as the least squares estimator. In fact, one can show that $\hat{\beta}_i$ is the UMVUE of β_i, $i = 1, 2, \ldots, k$, by the usual methods.

Example 3. Suppose that a random variable Y is linearly related to a mathematical variable x that is not random (see Example 2). Let $Y_1, Y_2, \ldots, Y_n$ be observations made at different known values $x_1, x_2, \ldots, x_n$ of x. For example, $x_1, x_2, \ldots, x_n$ may represent different levels of fertilizer, and $Y_1, Y_2, \ldots, Y_n$, respectively, the corresponding yields of a crop. Also, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ represent unobservable RVs that may be errors of measurements. Then

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, 2, \dots, n,$$

and we wish to test whether $\beta_1 = 0$, that the fertilizer levels do not affect the yield. Here

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix},$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1)', \quad \text{and} \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'.$$

The hypothesis to be tested is $H_0: \beta_1 = 0$, so that with $\mathbf{H} = (0, 1)$, the null hypothesis can be written as $H_0: \mathbf{H}\boldsymbol{\beta} = 0$. This is a problem of *linear regression.*

Similarly, we may assume that the regression of Y on x is *quadratic:*

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon,$$

and we may wish to test that a linear function will be sufficient to describe the relationship, that is, $\beta_2 = 0$. Here $\mathbf{X}$ is the $n \times 3$ matrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix},$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)', \quad \text{and} \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)',$$

and $\mathbf{H}$ is the 1×3 matrix (0, 0, 1).

In another example of regression, the Y's can be written as

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon,$$

and we wish to test the hypothesis that $\beta_1 = \beta_2 = \beta_3$. In this case, $\mathbf{X}$ is the matrix

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} \end{pmatrix},$$

and $\mathbf{H}$ may be chosen to be the 2×3 matrix

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}.$$

Example 4. Another important example of the general linear hypothesis involves the *analysis of variance*. We have already derived tests of hypotheses regarding the equality of the means of two normal populations when the variances are equal. In practice, one is frequently interested in the equality of several means when the variances are the same, that is, one has k samples from $\mathcal{N}(\mu_1, \sigma^2), \dots, \mathcal{N}(\mu_k, \sigma^2)$, where σ^2 is unknown and one wants to test $H_0\colon \mu_1 = \mu_2 = \cdots = \mu_k$ (see Example 1). Such a situation is of common occurrence in agricultural experiments. Suppose that k treatments are applied to experimental units (plots), the ith treatment is applied to n_i randomly chosen units, $i = 1, 2, \dots, k$, $\sum_{i=1}^{k} n_i = n$, and the observation y_{ij} represents some numerical characteristic (yield) of the jth experimental unit under the ith treatment. Suppose also that

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad j = 1, 2, \dots, n; \quad i = 1, 2, \dots, k,$$

where ε_{ij} are iid $\mathcal{N}(0, \sigma^2)$ RVs. We are interested in testing $H_0\colon \mu_1 = \mu_2 = \cdots = \mu_k$. We write

$$\mathbf{Y} = (Y_{11}, Y_{12}, \dots, Y_{1n_1}, Y_{21}, Y_{22}, \dots, Y_{2n_2}, \dots, Y_{k_1}, Y_{k_2}, \dots, Y_{kn_k})',$$
$$\boldsymbol{\beta} = (\mu_1, \mu_2, \dots, \mu_k)',$$

and

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_k} \end{pmatrix},$$

where $\mathbf{1}_{n_i} = (1, 1, \dots, 1)'$ is the n_i-vector ($i = 1, 2, \dots, k$), each of whose elements is unity. Thus $\mathbf{X}$ is $n \times k$. We can choose

$$\mathbf{H} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}$$

so that $H_0\colon \mu_1 = \mu_2 = \cdots = \mu_k$ is of the form $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$. Here $\mathbf{H}$ is a $(k-1) \times k$ matrix.

The model described in this example is frequently referred to as a *one-way analysis of variance model*. This is a very simple example of an analysis of variance model. Note that the matrix $\mathbf{X}$ is of a very special type; namely, the elements of $\mathbf{X}$ are either 0 or 1. $\mathbf{X}$ is known as a *design matrix*.

Returning to our general model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

we wish to test the null hypothesis H_0: $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$. We will compute the generalized likelihood ratio test and the distribution of the test statistic. To do so, we assume that $\boldsymbol{\varepsilon}$ has a multivariate normal distribution with mean vector $\mathbf{0}$ and variance–covariance matrix $\sigma^2\mathbf{I}_n$, where σ^2 is unknown and $\mathbf{I}_n$ is the $n \times n$ identity matrix. This means that $\mathbf{Y}$ has an n-variate normal distribution with mean $\mathbf{X}\boldsymbol{\beta}$ and dispersion matrix $\sigma^2\mathbf{I}_n$ for some $\boldsymbol{\beta}$ and some σ^2, both unknown. Here the parameter space Θ is the set of $(k+1)$-tuples $(\boldsymbol{\beta}', \sigma^2) = (\beta_1, \beta_2, \ldots, \beta_k, \sigma^2)$, and the joint PDF of the X's is given by

$$f_{\boldsymbol{\beta},\sigma^2}(y_1, y_2, \ldots, y_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2\right]$$

$$(8) \qquad = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right].$$

Theorem 1. Consider the linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{X}$ is an $n \times k$ matrix, (x_{ij}), $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, k$, of known constants and full rank $k < n$, $\boldsymbol{\beta}$ is a vector of unknown parameters $\beta_1, \beta_2, \ldots, \beta_k$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is a vector of nonobservable independent normal RVs with common variance σ^2 and mean $E\boldsymbol{\varepsilon} = \mathbf{0}$. The GLR test for testing the linear hypothesis H_0: $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{H}$ is an $r \times k$ matrix of full rank $r \le k$, is to reject H_0 at level α if $F \ge F_\alpha$, where $P_{H_0}\{F \ge F_\alpha\} = \alpha$ and F is the RV given by

$$(9) \qquad F = \frac{(\mathbf{Y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}})'(\mathbf{Y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}) - (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}.$$

In (9), $\hat{\boldsymbol{\beta}}$, and $\hat{\hat{\boldsymbol{\beta}}}$ are the MLEs of $\boldsymbol{\beta}$ under Θ and Θ_0, respectively. Moreover, the RV $[(n-k)/r]F$ has the F-distribution with $(r, n-k)$ d.f. under H_0.

Proof. The GLR test of H_0: $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$ is to reject H_0 if and only if $\lambda(\mathbf{y}) < c$, where

$$(10) \qquad \lambda(\mathbf{y}) = \frac{\sup_{\boldsymbol{\theta}\in\Theta_0} f_{\boldsymbol{\beta},\sigma^2}(\mathbf{y})}{\sup_{\boldsymbol{\theta}\in\Theta} f_{\boldsymbol{\beta},\sigma^2}(\mathbf{y})},$$

$\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$, and $\Theta_0 = \{(\boldsymbol{\beta}', \sigma^2)' \colon \mathbf{H}\boldsymbol{\beta} = \mathbf{0}\}$. Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\sigma}^2)'$ be the MLE of $\boldsymbol{\theta}' \in \Theta$, and $\hat{\hat{\boldsymbol{\theta}}} = (\hat{\hat{\boldsymbol{\beta}}}', \hat{\hat{\sigma}}^2)'$ be the MLE of $\boldsymbol{\theta}$ under H_0, that is, when $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$. It is easily seen that $\hat{\boldsymbol{\beta}}$ is the value of $\boldsymbol{\beta}$ that minimizes $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, and

$$(11) \qquad \hat{\sigma}^2 = n^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

Similarly, $\hat{\hat{\boldsymbol{\beta}}}$ is the value of $\boldsymbol{\beta}$ that minimizes $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ subject to $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$,

and

$$\hat{\hat{\sigma}}^2 = n^{-1}(\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}})'(\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}). \tag{12}$$

It follows that

$$\lambda(\mathbf{y}) = \left(\frac{\hat{\sigma}^2}{\hat{\hat{\sigma}}^2}\right)^{n/2}. \tag{13}$$

The critical region $\lambda(\mathbf{y}) < c$ is equivalent to the region $\{\lambda(\mathbf{y})\}^{-2/n} < \{c\}^{-2/n}$, which is of the form

$$\frac{\hat{\hat{\sigma}}^2}{\hat{\sigma}^2} > c_1. \tag{14}$$

This may be written as

$$\frac{(\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}})'(\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}})}{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})} > c_1 \tag{15}$$

or, equivalently, as

$$\frac{(\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}})'(\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}) - (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})} > c_1 - 1. \tag{16}$$

It remains to determine the distribution of the test statistic. For this purpose it is convenient to reduce the problem to the *canonical form*. Let V_n be the vector space of the observation vector $\mathbf{Y}$, V_k be the subspace of V_n generated by the column vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ of $\mathbf{X}$, and V_{k-r} be the subspace of V_k in which $E\mathbf{Y}$ is postulated to lie under H_0. We change variables from $Y_1, Y_2, \ldots, Y_n$ to $Z_1, Z_2, \ldots, Z_n$, where $Z_1, Z_2, \ldots, Z_n$ are independent normal RVs with common variance σ^2 and means $EZ_i = \theta_i$, $i = 1, 2, \ldots, k$, $EZ_i = 0$, $i = k+1, \ldots, n$. This is done as follows. Let us choose an orthonormal basis of $k - r$ column vectors $\{\boldsymbol{\alpha}_i\}$ for V_{k-r}, say $\{\boldsymbol{\alpha}_{r+1}, \boldsymbol{\alpha}_{r+2}, \ldots, \boldsymbol{\alpha}_k\}$. We extend this to an orthonormal basis $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+1}, \ldots, \boldsymbol{\alpha}_k\}$ for V_k, and then extend once again to an orthonormal basis $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_k, \boldsymbol{\alpha}_{k+1}, \ldots, \boldsymbol{\alpha}_n\}$ for V_n. This is always possible.

Let $z_1, z_2, \ldots, z_n$ be the coordinates of $\mathbf{y}$ relative to the basis $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n\}$. Then $z_i = \boldsymbol{\alpha}_i'\mathbf{y}$ and $\mathbf{z} = \mathbf{PY}$, where $\mathbf{P}$ is an orthogonal matrix with ith row $\boldsymbol{\alpha}_i'$. Thus $EZ_i = E\boldsymbol{\alpha}_i'\mathbf{Y} = \boldsymbol{\alpha}_i'\mathbf{X}\boldsymbol{\beta}$, and $E\mathbf{Z} = \mathbf{PX}\boldsymbol{\beta}$. Since $\mathbf{X}\boldsymbol{\beta} \in V_k$ (Remark 1), it follows that $\boldsymbol{\alpha}_i'\mathbf{X}\boldsymbol{\beta} = 0$ for $i > k$. Similarly, under H_0, $\mathbf{X}\boldsymbol{\beta} \in V_{k-r} \subset V_k$, so that $\boldsymbol{\alpha}_i'\mathbf{X}\boldsymbol{\beta} = 0$ for $i \le r$. Let us write $\boldsymbol{\omega} = \mathbf{PX}\boldsymbol{\beta}$. Then $\omega_{k+1} = \omega_{k+2} = \cdots = \omega_n = 0$, and under H_0, $\omega_1 = \omega_2 = \cdots = \omega_r = 0$. Finally, from Corollary 2 of Theorem 5.4.6 it follows that $Z_1, Z_2, \ldots, Z_n$ are independent normal RVs with the same variance σ^2 and $EZ_i = \omega_i$, $i = 1, 2, \ldots, n$. We have thus transformed the problem to the following simpler canonical form:

$$
(17) \qquad \begin{cases} \Omega: & Z_i \text{ are independent } \mathcal{N}(\omega_i, \sigma^2), \quad i = 1, 2, \ldots, n, \\ & \omega_{k+1} = \omega_{k+2} = \cdots = \omega_n = 0, \\ H_0: & \omega_1 = \omega_2 = \cdots = \omega_r = 0. \end{cases}
$$

Now

$$
\begin{aligned}
(18) \qquad (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= (\mathbf{P}'\mathbf{z} - \mathbf{P}'\boldsymbol{\omega})'(\mathbf{P}'\mathbf{z} - \mathbf{P}'\boldsymbol{\omega}) \\
&= (\mathbf{z} - \boldsymbol{\omega})'(\mathbf{z} - \boldsymbol{\omega}) \\
&= \sum_{i=1}^{k} (z_i - \omega_i)^2 + \sum_{i=k+1}^{n} z_i^2.
\end{aligned}
$$

The quantity $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ is minimized if we choose $\hat{\omega}_i = z_i$, $i = 1, 2, \ldots, k$, so that

$$
(19) \qquad (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \sum_{i=k+1}^{n} z_i^2.
$$

Under $H_0, \omega_1 = \omega_2 = \cdots = \omega_r = 0$, so that $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ will be minimized if we choose $\hat{\hat{\omega}}_i = z_i$, $i = r + 1, \ldots, k$. Thus

$$
(20) \qquad (\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}})'(\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}) = \sum_{i=1}^{r} z_i^2 + \sum_{i=k+1}^{n} z_i^2.
$$

It follows that

$$
F = \frac{\sum_{i=1}^{r} Z_i^2}{\sum_{i=k+1}^{n} Z_i^2}.
$$

Now $\sum_{i=k+1}^{n} Z_i^2/\sigma^2$ has a $\chi^2(n-k)$ distribution, and under H_0, $\sum_{i=1}^{r} Z_i^2/\sigma^2$ has a $\chi^2(r)$ distribution. Since $\sum_{i=1}^{r} Z_i^2$ and $\sum_{i=k+1}^{n} Z_i^2$ are independent, we see that $[(n-k)/r]F$ is distributed as $F(r, n-k)$ under H_0, as asserted. This completes the proof of the theorem.

Remark 4. In practice, one does not need to find a transformation that reduces the problem to the canonical form. As will be done in the following sections, one simply computes the estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\hat{\boldsymbol{\theta}}}$ and then computes the test statistic in any of the equivalent forms (14), (15), or (16) to apply the F-test.

Remark 5. The computation of $\hat{\boldsymbol{\beta}}$, $\hat{\hat{\boldsymbol{\beta}}}$ is greatly facilitated, in view of Remark 3, by using the principle of least squares. Indeed, this was done in the proof of Theorem 1 when we reduced the problem of maximum likelihood estimation to that of minimization of sum of squares $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

Remark 6. The distribution of the test statistic under H_1 is easily determined. We note that $Z_i/\sigma \sim \mathcal{N}(\omega_i/\sigma, 1)$ for $i = 1, 2, \ldots, r$, so that $\sum_{i=1}^{r} Z_i^2/\sigma^2$ has a noncentral chi-square distribution with r d.f. and noncentrality parameter $\delta = \sum_{i=1}^{r} \omega_i^2/\sigma^2$. It follows that $[(n-k)/r]F$ has a noncentral F-distribution with d.f. $(r, n-k)$ and noncentrality parameter δ. Under $H_0, \delta = 0$, so that $[(n-k)/r]F$ has a central $F(r, n-k)$ distribution. Since $\sum_{i=1}^{r} \omega_i^2 = \sum_{i=1}^{r}(EZ_i)^2$, it follows from (19) and (20) that if we replace each observation Y_i by its expected value in the numerator of (16), we get $\sigma^2\delta$.

Remark 7. The general linear hypothesis makes use of the assumption of common variance. For instance, in Example 4, $Y_{ij} \sim \mathcal{N}(\mu_i, \sigma^2)$, $j = 1, 2, \ldots, k$. Let us suppose that $Y_{ij} \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, k$. Then we need to test that $\sigma_1 = \sigma_2 = \cdots = \sigma_k$ before we can apply Theorem 1. The case $k = 2$ has already been considered in Section 10.3. For the case where $k > 2$ one can show that a UMP unbiased test does not exist. A large-sample approximation is described by Lehmann [62, pp. 376–377]. It is beyond the scope of this book to consider the effects of departures from the underlying assumptions. We refer the reader to Scheffé [99, Chap. 10], for a discussion of this topic.

PROBLEMS 12.2

1. Show that any solution of the normal equations (5) minimizes the sum of squares $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$.
2. Show that the least squares estimator given in (6) is an unbiased estimator of $\boldsymbol{\beta}$. If the RVs Y_i are uncorrelated with common variance σ^2, show that the covariance matrix of the $\hat{\beta}_i$'s is given by (7).
3. Under the assumption that $\boldsymbol{\varepsilon}$ [in model (2)] has a multivariate normal distribution with mean 0 and dispersion matrix $\sigma^2 I_n$, show that the least squares estimators and the MLEs of $\boldsymbol{\beta}$ coincide.
4. Prove statements (11) and (12).
5. Determine the expression for the least squares estimator of $\boldsymbol{\beta}$ subject to $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$.

12.3 REGRESSION MODEL

In this section we consider a simple *linear regression model* as a special case of the general linear hypothesis and show how some inferential questions about the parameters of the regression equation can be answered. Let $x_1, x_2, \ldots, x_n$ be n given numbers, and suppose that

$$(1) \qquad Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, 2, \dots, n,$$

where β_0, β_1 are unknown parameters and ε_i are independent normal RVs with $E\varepsilon_i = 0$ and $\text{var}(\varepsilon_i) = \sigma^2$, $i = 1, 2, \dots, n$. Also, σ^2 is assumed to be unknown. Our object is to test hypotheses concerning β_0 and β_1 and to construct confidence intervals for β_0 and β_1. Rewriting (1) in the usual fashion, we have

$$(2) \qquad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\boldsymbol{\beta} = (\beta_0, \beta_1)' \quad \text{and} \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

Clearly, $Y_1, Y_2, \dots, Y_n$ are independent normal RVs with $EY_i = \beta_0 + \beta_1 x_i$ and $\text{var}(Y_i) = \sigma^2$, $i = 1, 2, \dots, n$, and $\mathbf{Y}$ is an n-variate normal random vector with mean $\mathbf{X}\boldsymbol{\beta}$ and variance $\sigma^2 \mathbf{I}_n$. The joint PDF of $\mathbf{Y}$ is given by

$$(3) \qquad f(\mathbf{y}; \beta_0, \beta_1, \sigma^2) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \right].$$

It easily follows that the MLEs for β_0, β_1, and σ^2 are given by

$$(4) \qquad \hat{\beta}_0 = \frac{\sum_{i=1}^{n} Y_i}{n} - \hat{\beta}_1 \overline{x},$$

$$(5) \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2},$$

and

$$(6) \qquad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2,$$

where $\overline{x} = n^{-1} \sum_{i=1}^{n} x_i$.

If we wish to test $H_0\colon \beta_1 = 0$, we take $\mathbf{H} = (0, 1)$, so that the model is a special case of the general linear hypothesis with $k = 2$, $r = 1$. Under H_0 the MLEs are

$$(7) \qquad \hat{\hat{\beta}}_0 = \overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

and

(8) $$\hat{\hat{\sigma}}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

Thus

(9) $$F = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 - \sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2}$$
$$= \frac{\hat{\beta}_1^2\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_1)^2}.$$

From Theorem 12.2.1, the statistic $[(n-2)/1]F$ has a central $F(1, n-2)$ distribution under H_0. Since $F(1, n-2)$ is the square of a $t(n-2)$, the likelihood ratio test rejects H_0 if

(10) $$|\hat{\beta}_1|\left[\frac{(n-2)\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2}\right]^{1/2} > c_0,$$

where c_0 is determined from t-tables for $n-2$ d.f.

For testing $H_0\colon \beta_0 = 0$, we choose $\mathbf{H} = (1, 0)$ so that the model is again a special case of the general linear hypothesis. In this case

$$\hat{\hat{\beta}}_1 = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$$

and

(11) $$\hat{\hat{\sigma}}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\hat{\beta}}_1 x_i)^2.$$

It follows that

(12) $$F = \frac{\sum_{i=1}^{n}(Y_i - \hat{\hat{\beta}}_1 x_i)^2 - \sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2},$$

and since

(13) $$\hat{\hat{\beta}}_1 = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y}) + n\overline{x}\overline{Y}}{\sum_{i=1}^{n} x_i^2}$$
$$= \frac{\hat{\beta}_1\sum_{i=1}^{n}(x_i - \overline{x})^2 + n\overline{x}(\hat{\beta}_0 + \hat{\beta}_1\overline{x})}{\sum_{i=1}^{n} x_i^2}$$
$$= \hat{\beta}_1 + \frac{n\hat{\beta}_0\overline{x}}{\sum_{i=1}^{n} x_i^2},$$

we can write the numerator of F as

$$(14)\quad \sum_{i=1}^{n}(Y_i - \hat{\hat{\beta}}_1 x_i)^2 - \sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2$$

$$= \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_1 x_i + \hat{\beta}_1\overline{x} - \overline{Y} + \overline{Y} - \hat{\beta}_1\overline{x} - \frac{n\hat{\beta}_0\overline{x}x_i}{\sum_{i=1}^{n} x^2}\right)^2$$

$$- \sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2$$

$$= \sum_{i=1}^{n}\left(\overline{Y} - \hat{\beta}_1\overline{x} - \frac{n\hat{\beta}_0\overline{x}x_i}{\sum_{i=1}^{n} x_i^2}\right)^2 + 2\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 x_i + \hat{\beta}_1\overline{x} - \overline{Y})$$

$$\cdot \left(\overline{Y} - \hat{\beta}_1\overline{x} - \frac{n\hat{\beta}_0\overline{x}x_i}{\sum_{i=1}^{n} x_i^2}\right)$$

$$= \frac{\hat{\beta}_0^2 n \sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{i=1}^{n} x_i^2}.$$

It follows from Theorem 12.2.1 that the statistic

$$(15)\quad \frac{\hat{\beta}_0\sqrt{n\sum_{i=1}^{n}(x_i - \overline{x})^2 / \sum_{i=1}^{n} x_i^2}}{\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2/(n-2)}}$$

has a central t-distribution with $n-2$ d.f. under $H_0\colon \beta_0 = 0$. The rejection region is therefore given by

$$(16)\quad \frac{|\hat{\beta}_0|\sqrt{n\sum_{i=1}^{n}(x_i - \overline{x})^2 / \sum_{i=1}^{n} x_i^2}}{\sqrt{\sum_{i=1}^{n}(Y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2/(n-2)}} > c_0,$$

where c_0 is determined from the tables of $t(n-2)$ distribution for a given level of significance α.

For testing $H_0\colon \beta_0 = \beta_1 = 0$, we choose $\mathbf{H} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$, so that the model is again a special case of the general linear hypothesis with $r = 2$. In this case

$$(17)\quad \hat{\hat{\sigma}}^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

and

$$F = \frac{\sum_{i=1}^{n} Y_i^2 - \sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta} x_i)^2}
= \frac{n\overline{Y}^2 + \hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}
= \frac{n(\hat{\beta}_0 + \hat{\beta}_1\overline{x})^2 + \hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}. \tag{18}$$

From Theorem 12.2.1, the statistic $[(n-2)/2]F$ has a central $F(2, n-2)$ distribution under $H_0\colon \beta_0 = \beta_1 = 0$. It follows that the level-α rejection region for H_0 is given by

$$\frac{n-2}{2} F > c_0, \tag{19}$$

where F is given by (18) and c_0 is the upper α percent point under the $F(2, n-2)$ distribution.

Remark 1. It is quite easy to modify the analysis above to obtain tests of null hypotheses $\beta_0 = \beta_0'$, $\beta_1 = \beta_1'$, and $(\beta_0, \beta_1)' = (\beta_0', \beta_1')'$, where β_0', β_1' are given real numbers (Problem 4).

Remark 2. The confidence intervals for β_0, β_1 are also easily obtained. One can show that a $(1-\alpha)$-level confidence interval for β_0 is given by

$$\left(\hat{\beta}_0 - t_{n-2,\alpha/2}\sqrt{\frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n(n-2)\sum_{i=1}^{n}(x_i - \overline{x})^2}},\right.
\left.\hat{\beta}_0 + t_{n-2,\alpha/2}\sqrt{\frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n(n-2)\sum_{i=1}^{n}(x_i - \overline{x})^2}}\right), \tag{20}$$

and that for β_1 is given by

$$\left(\hat{\beta}_1 - t_{n-2,\alpha/2}\sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{(n-2)\sum_{i=1}^{n}(x_i - \overline{x})^2}},\right.
\left.\hat{\beta}_1 + t_{n-2,\alpha/2}\sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{(n-2)\sum_{i=1}^{n}(x_i - \overline{x})^2}}\right). \tag{21}$$

Similarly, one can obtain confidence sets for $(\beta_0, \beta_1)'$ from the likelihood ratio test of $(\beta_0, \beta_1)' = (\beta_0', \beta_1')'$. It can be shown that the collection of sets of points $(\beta_0, \beta_1)'$ satisfying

$$(22)\qquad \frac{(n-2)[n(\hat\beta_0-\beta_0)^2+2n\overline{x}(\hat\beta_0-\beta_0)(\hat\beta_1-\beta_1)+\sum_{i=1}^n x_i^2(\hat\beta_1-\beta_1)^2]}{2\sum_{i=1}^n(Y_i-\hat\beta_0-\hat\beta_1 x_i)^2}$$
$$\leq F_{2,n-2,\alpha}$$

is a $(1-\alpha)$-level collection of confidence sets (ellipsoids) for $(\beta_0, \beta_1)'$ centered at $(\hat\beta_0, \hat\beta_1)'$.

Remark 3. Sometimes interest lies in constructing a confidence interval on the unknown linear regression function $E\{Y \mid x_0\} = \beta_0 + \beta_1 x_0$ for a given value of x, or on a value of Y given $x = x_0$. We assume that x_0 is a value of x distinct from $x_1, x_2, \ldots, x_n$. Clearly, $\hat\beta_0 + \hat\beta_1 x_0$ is the maximum likelihood estimator of $\beta_0 + \beta_1 x_0$. This is also the best linear unbiased estimator. Let us write $\hat E\{Y \mid x_0\} = \hat\beta_0 + \hat\beta_1 x_0$. Then

$$\begin{aligned}\hat E\{Y \mid x_0\} &= \overline{Y} - \hat\beta_1\overline{x} + \hat\beta_1 x_0\\ &= \overline{Y} + (x_0-\overline{x})\frac{\sum_{i=1}^n (x_i-\overline{x})(Y_i-\overline{Y})}{\sum_{i=1}^n (x_i-\overline{x})^2},\end{aligned}$$

which is clearly a linear function of normal RVs Y_i. It follows that $\hat E\{Y \mid x_0\}$ is also normally distributed with mean $E(\hat\beta_0 + \hat\beta_1 x_0) = \beta_0 + \beta_1 x_0$ and variance

$$(23)\qquad \begin{aligned}\operatorname{var}(\hat E\{Y \mid x_0\}) &= E(\hat\beta_0 - \beta_0 + \hat\beta_1 x_0 - \beta_1 x_0)^2\\ &= \operatorname{var}(\hat\beta_0) + x_0^2\operatorname{var}(\hat\beta_1) + 2x_0\operatorname{cov}(\hat\beta_0, \hat\beta_1)\\ &= \sigma^2\left[\frac{1}{n} + \frac{(\overline{x}-x_0)^2}{\sum_{i=1}^n (x_i-\overline{x})^2}\right],\end{aligned}$$

(see Problem 6). It follows that

$$(24)\qquad \frac{\hat\beta_0 + \hat\beta_1 x_0 - \beta_0 - \beta_1 x_0}{\sigma\{(1/n) + [(\overline{x}-x_0)^2/\sum_{i=1}^n (x_i-\overline{x})^2]\}^{1/2}}$$

is $\mathcal{N}(0, 1)$. But σ is not known, so that we cannot use (24) to construct a confidence interval for $E\{Y \mid x_0\}$. Since $n\hat\sigma^2/\sigma^2$ is a $\chi^2(n-2)$ RV and $n\hat\sigma^2/\sigma^2$ is independent of $\hat\beta_0 + \hat\beta_1 x_0$ (why?), it follows that

$$(25)\qquad \sqrt{n-2}\,\frac{\hat\beta_0 + \hat\beta_1 t_0 - \beta_0 - \beta_1 x_0}{\hat\sigma\{1 + n[(\overline{x}-x_0)^2/\sum_{i=1}^n (x_i-\overline{x})^2]\}^{1/2}}$$

has a $t(n-2)$ distribution. Thus a $(1-\alpha)$-level confidence interval for $\beta_0 + \beta_1 x_0$ is given by

$$(26)\qquad \left(\hat{\beta}_0+\hat{\beta}_1x_0-t_{n-2,\alpha/2}\,\hat{\sigma}\sqrt{\frac{n}{n-2}\left[\frac{1}{n}+\frac{(\overline{x}-x_0)^2}{\sum_{i=1}^n(x_i-\overline{x})^2}\right]},\right.$$
$$\left.\hat{\beta}_0+\hat{\beta}_1x_0+t_{n-2,\alpha/2}\,\hat{\sigma}\sqrt{\frac{n}{n-2}\left[\frac{1}{n}+\frac{(\overline{x}-x_0)^2}{\sum_{i=1}^n(x_i-\overline{x})^2}\right]}\right).$$

In a similar manner, one can show (Problem 7) that

$$(27)\qquad \left(\hat{\beta}_0+\hat{\beta}_1x_0-t_{n-2,\alpha/2}\,\hat{\sigma}\sqrt{\frac{n}{n-2}\left[\frac{n+1}{n}+\frac{(\overline{x}-x_0)^2}{\sum_{i=1}^n(x_i-\overline{x})^2}\right]},\right.$$
$$\left.\hat{\beta}_0+\hat{\beta}_1x_0+t_{n-2,\alpha/2}\,\hat{\sigma}\sqrt{\frac{n}{n-2}\left[\frac{n+1}{n}+\frac{(\overline{x}-x_0)^2}{\sum_{i=1}^n(x_i-\overline{x})^2}\right]}\right)$$

is a $(1-\alpha)$-level confidence interval for $Y_0=\beta_0+\beta_1x_0+\varepsilon$, that is, for the estimated value Y_0 of Y at x_0.

Remark 4. The simple regression model (2) considered above can be generalized in many directions. Thus we may consider EY as a polynomial in x of a degree higher than 1, or we may regard EY as a function of several variables. Some of these generalizations will be taken up in the problems.

Remark 5. Let $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ be a sample from a bivariate normal population with parameters $EX=\mu_1$, $EY=\mu_2$, $\text{var}(X)=\sigma_1^2$, $\text{var}(Y)=\sigma_2^2$, and $\text{cov}(X,Y)=\rho$. In Section 7.7 we computed the PDF of the sample correlation coefficient R and showed (Remark 7.7.4) that the statistic

$$(28)\qquad T=R\sqrt{\frac{n-2}{1-R^2}}$$

has a $t(n-2)$ distribution, provided that $\rho=0$. If we wish to test $\rho=0$, that is, the independence of two jointly distributed normal RVs, we can base a test on the statistic T. Essentially, we are testing that the population covariance is 0, which implies that the population regression coefficients are 0. Thus we are testing, in particular, that $\beta_1=0$. It is therefore not surprising that (28) is identical with (10). We emphasize that we derived (28) for a bivariate normal population, but (10) was derived by taking the X's as fixed and the distribution of Y's as normal. Note that for a bivariate normal population, $E\{Y\mid x\}=\mu_2+\rho(\sigma_2/\sigma_1)(x-\mu_1)$ is linear, consistent with our model (1) or (2).

Example 1. Let us assume that the following data satisfy a linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

x	0	1	2	3	4	5
y	0.475	1.007	0.838	−0.618	1.378	0.943

Let us test the null hypothesis that $\beta_1 = 0$. We have

$$\bar{x} = 2.5, \qquad \sum_{i=0}^{5}(x_i - \bar{x})^2 = 17.5, \qquad \bar{y} = 0.671,$$

$$\sum_{i=0}^{5}(x_i - \bar{x})(y_i - \bar{y}) = 0.9985,$$

$$\hat{\beta}_1 = 0.0571, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 0.5279,$$

$$\sum_{i=0}^{5}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 2.3571,$$

and

$$|\hat{\beta}_1|\sqrt{\frac{(n-2)\sum(x_i - \bar{x})^2}{\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}} = 0.3106.$$

Since $t_{n-2,\alpha/2} = t_{4,0.025} = 2.776 > 0.3106$, we accept H_0 at level $\alpha = 0.05$.

Let us next find a 95 percent confidence interval for $E\{Y \mid x = 7\}$. This is given by (26). We have

$$t_{n-2,\alpha/2}\hat{\sigma}\sqrt{\frac{n}{n-2}\left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum(x_i - \bar{x})^2}\right]} = 2.776\sqrt{\frac{2.3571}{6}}\sqrt{\frac{6}{4}\left(\frac{1}{6} + \frac{20.25}{17.5}\right)}$$
$$= 2.3707,$$
$$\hat{\beta}_0 + \hat{\beta}_1 x_0 = 0.5279 + 0.0571 \times 7$$
$$= 0.9276,$$

so that the 95 percent confidence interval is $(-1.4431, 3.2983)$.

(The data were produced from Table ST6, random numbers with $\mu = 0$, $\sigma = 1$, by letting $\beta_0 = 1$ and $\beta_1 = 0$ so that $E\{Y \mid x\} = \beta_0 + \beta_1 x = 1$, which surely lies in the interval.)

PROBLEMS 12.3

1. Prove statements (4), (5), and (6).
2. Prove statements (7) and (8).
3. Prove statement (11).

4. Obtain tests of null hypotheses $\beta_0 = \beta_0'$, $\beta_1 = \beta_1'$, and $(\beta_0, \beta_1)' = (\beta_0', \beta_1')'$, where β_0', β_1' are given real numbers.
5. Obtain the confidence intervals for β_0 and β_1 as given in (20) and (21), respectively.
6. Derive the expression for $\text{var}(\hat{E}\{Y \mid x_0\})$ as given in (24).
7. Show that the interval given in (27) is a $(1-\alpha)$-level confidence interval for $Y_0 = \beta_0 + \beta_1 x_0 + \varepsilon$, the estimated value of Y at x_0.
8. Suppose that the regression of Y on the (mathematical) variable x is a quadratic

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i,$$

where $\beta_0, \beta_1, \beta_2$ are unknown parameters, $x_1, x_2, \ldots, x_n$ are known values of x, and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are unobservable RVs that are assumed to be independently normally distributed with common mean 0 and common variance σ^2 (see Example 12.2.3). Assume that the coefficient vectors $(x_1^k, x_2^k, \ldots, x_n^k)$, $k = 0, 1, 2$, are linearly independent. Write the normal equations for estimating the β's and derive the generalized likelihood ratio test of $\beta_2 = 0$.
9. Suppose that the Y's can be written as

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i,$$

where x_{i1}, x_{i2}, x_{i3} are three mathematical variables, and ε_i are iid $\mathcal{N}(0, 1)$ RVs. Assuming that the matrix $\mathbf{X}$ (see Example 12.2.3) is of full rank, write the normal equations and derive the likelihood ratio test of the null hypothesis $H_0\colon \beta_1 = \beta_2 = \beta_3$.
10. The following table gives the weight Y (grams) of a crystal suspended in a saturated solution against the time suspended T (days).

Time, T	0	1	2	3	4	5	6
Weight, Y	0.4	0.7	1.1	1.6	1.9	2.3	2.6

(a) Find the linear regression line of Y on T.

(b) Test the hypothesis that $\beta_0 = 0$ in the linear regression model $Y_i = \beta_0 + \beta_1 T_i + \varepsilon_i$.

(c) Obtain a 0.95 level confidence interval for β_0.

12.4 ONE-WAY ANALYSIS OF VARIANCE

In this section we return to the problem of one-way analysis of variance considered in Examples 12.2.1 and 12.2.4. Consider the model

$$(1) \qquad Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad j = 1, 2, \ldots, n_i; \qquad i = 1, 2, \ldots, k,$$

as described in Example 12.2.4. In matrix notation we write

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where

$$\mathbf{Y} = (Y_{11}, Y_{12}, \ldots, Y_{1n_1}, Y_{21}, Y_{22}, \ldots, Y_{2n_2}, \ldots, Y_{k1}, Y_{k2}, \ldots, Y_{kn_k})',$$

$$\boldsymbol{\beta} = (\mu_1, \mu_2, \ldots, \mu_k)',$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_k} \end{pmatrix},$$

and

$$\boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{1n_1}, \varepsilon_{21}, \varepsilon_{22}, \ldots, \varepsilon_{2n_2}, \ldots, \varepsilon_{k1}, \varepsilon_{k2}, \ldots, \varepsilon_{kn_k})'.$$

As in Example 12.2.4, $\mathbf{Y}$ is a vector of n-observations ($n = \sum_{i=1}^{k} n_i$), whose components Y_{ij} are subject to random error $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, $\boldsymbol{\beta}$ is a vector of k unknown parameters, and $\mathbf{X}$ is a design matrix. We wish to find a test of H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$ against all alternatives. We may write H_0 in the form $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{H}$ is a $(k-1) \times k$ matrix of rank $(k-1)$, which can be chosen to be

$$\mathbf{H} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}.$$

Let us write $\mu_1 = \mu_2 = \cdots \mu_k = \mu$ under H_0. The joint PDF of $\mathbf{Y}$ is given by

$$f(\mathbf{y}; \mu_1, \mu_2, \ldots, \mu_k, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \mu_i)^2\right], \tag{3}$$

and under H_0 by

$$f(\mathbf{x}; \mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \mu)^2\right]. \tag{4}$$

It is easy to check that the MLEs are

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} = \overline{y}_{i\cdot}, \qquad i = 1, 2, \ldots, k, \tag{5}$$

(6) $$\hat{\sigma}^2 = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i.)^2}{n},$$

(7) $$\hat{\hat{\mu}} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}}{n} = \overline{y},$$

and

(8) $$\hat{\hat{\sigma}}^2 = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \overline{y})^2}{n}.$$

By Theorem 12.2.1, the likelihood ratio test is to reject H_0 if

(9) $$\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y})^2 - \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i.)^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i.)^2} \frac{n-k}{k-1} \geq F_0,$$

where F_0 is the upper α percent point in the $F(k-1, n-k)$ distribution. Since

(10) $$\begin{aligned}\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y})^2 &= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i. + \overline{Y}_i. - \overline{Y})^2 \\ &= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i.)^2 + \sum_{i=1}^k n_i (\overline{Y}_i. - \overline{Y})^2,\end{aligned}$$

we may rewrite (9) as

(11) $$\frac{\sum_{i=1}^k n_i (\overline{Y}_i. - \overline{Y})^2/(k-1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i.)^2/(n-k)} \geq F_0.$$

It is usual to call the sum of squares in the numerator of (11) the *between sum of squares* (BSS), and the sum of squares in the denominator of (11) the *within sum of squares* (WSS). The results are conveniently displayed in an *analysis of variance table* in the following form:

One-Way Analysis of Variance

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Sum of Squares	F-Ratio
Between	$\text{BSS} = \sum_{i=1}^k n_i (\overline{Y}_i. - \overline{Y})^2$	$k-1$	$\text{BSS}/(k-1)$	$\dfrac{\text{BSS}/(k-1)}{\text{WSS}/(n-k)}$
Within	$\text{WSS} = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i.)^2$	$n-k$	$\text{WSS}/(n-k)$	
Mean	$n\overline{Y}^2$	1		
Total	$\text{TSS} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2$	n		

The third row, "Mean," has been included to make the total of the second column add up to the total sum of squares (TSS), $\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2$.

Example 1. The lifetimes (in hours) of samples from three different brands of batteries, Y_1, Y_2, and Y_3, were recorded, with the following results:

Y_1	Y_2	Y_3
40	60	60
30	40	50
50	55	70
50	65	65
30		75
		40

We wish to test whether the three brands have different average lifetimes. We will assume that the three samples come from normal populations with common (unknown) standard deviation σ.

From the data $n_1 = 5$, $n_2 = 4$, $n_3 = 6$, $n = 15$, and

$$\bar{y}_1 = \frac{200}{5} = 40, \quad \bar{y}_2 = \frac{220}{4} = 55, \quad \bar{y}_3 = \frac{360}{6} = 60,$$

$$\sum_{i=1}^{5}(y_{1i} - \bar{y}_1)^2 = 400, \quad \sum_{i=1}^{4}(y_{2i} - \bar{y}_2)^2 = 350, \quad \sum_{i=1}^{6}(y_{3i} - \bar{y}_3)^2 = 850.$$

Also, the grand mean is

$$\bar{y} = \frac{200 + 220 + 360}{15} = \frac{780}{15} = 52.$$

Thus

$$\begin{aligned}\text{BSS} &= 5(40-52)^2 + 4(55-52)^2 + 6(60-52)^2 \\ &= 1140\end{aligned}$$

and

$$\text{WSS} = 400 + 350 + 850 = 1600.$$

Analysis of Variance

Source	SS	d.f.	MSS	F-Ratio
Between	1140	2	570	$570/133.33 = 4.28$
Within	1600	12	133.33	

Choosing $\alpha = 0.05$, we see that $F_0 = F_{2,12,0.05} = 3.89$. Thus we reject $H_0: \mu_1 = \mu_2 = \mu_3$ at level $\alpha = 0.05$.

Example 2. Three sections of the same elementary statistics course were taught by three instructors, I, II, and III. The final grades of students were recorded as follows:

I	II	III
95	88	68
33	78	79
48	91	91
76	51	71
89	85	87
82	77	68
60	31	79
77	62	16
	96	35
	81	

Let us test the hypothesis that the average grades given by the three instructors are the same at level $\alpha = 0.05$.

From the data $n_1 = 8$, $n_2 = 10$, $n_3 = 9$, $n = 27$, $\overline{y}_1 = 70$, $\overline{y}_2 = 74$, $\overline{y}_3 = 66$, $\sum_{i=1}^{8}(y_{1i} - \overline{y}_1)^2 = 3168$, $\sum_{i=1}^{10}(y_{2i} - \overline{y}_2)^2 = 3686$, $\sum_{i=1}^{9}(y_{3i} - \overline{y}_3)^2 = 4898$. Also, the grand mean is

$$\overline{y} = \frac{560 + 740 + 594}{27} = \frac{1894}{27} = 70.15.$$

Thus

$$\text{BSS} = 8(0.15)^2 + 10(3.85)^2 + 9(4.15)^2 = 303.4075$$

and

$$\text{WSS} = 3168 + 3686 + 4898 = 11{,}752.$$

Analysis of Variance

Source	SS	d.f.	MSS	F-Ratio
Between	303.41	2	151.70	151.70/489.67
Within	11,752.00	24	489.67	

We therefore cannot reject the null hypothesis that the average grades given by the three instructors are the same.

PROBLEMS 12.4

1. Prove statements (5), (6), (7), and (8).

2. The following are the coded values of the amounts of corn (in bushels per acre) obtained from four varieties, using unequal number of plots for the different varieties:

$$A: \quad 2, 1, 3, 2$$
$$B: \quad 3, 4, 2, 3, 4, 2$$
$$C: \quad 6, 4, 8$$
$$D: \quad 7, 6, 7, 4$$

Test whether there is a significant difference between the yields of the varieties.

3. A consumer interested in buying a new car has reduced his search to six different brands: D, F, G, P, V, T. He would like to buy the brand that gives the highest mileage per gallon of regular gasoline. One of his friends advises him that he should use some other method of selection, since the average mileages of the six brands are the same, and offers the following data in support of her assertion.

Distance Traveled (Miles) per Gallon of Gasoline

	Brand					
Car	D	F	G	P	V	T
1	42	38	28	32	30	25
2	35	33	32	36	35	32
3	37	28	35	27	25	24
4		37	37	26	30	
5				28	30	
6				19		

Should the consumer accept his friend's advice?

4. The following data give the ages of entering freshmen in independent random samples from three different universities, A, B, and C.

A	B	C
17	16	21
19	16	23
20	19	22
21		20
18		19

Test the hypothesis that the average ages of entering freshman at these universities are the same.

5. Five cigarette manufacturers claim that their product has low tar content. Independent random samples of cigarettes are taken from each manufacturer and the following tar levels (in milligrams) are recorded.

Brand	Tar Level (mg)
A	4.2, 4.8, 4.6, 4.0, 4.4
B	4.9, 4.8, 4.7, 5.0, 4.9, 5.2
C	5.4, 5.3, 5.4, 5.2, 5.5
D	5.8, 5.6, 5.5, 5.4, 5.6, 5.8
E	5.9, 6.2, 6.2, 6.8, 6.4, 6.3

Can the differences among the sample means be attributed to chance?

6. The quantity of oxygen dissolved in water is used as a measure of water pollution. Samples are taken at four locations in a lake and the quantity of dissolved oxygen is recorded as follows (lower reading corresponds to greater pollution):

Location	Quantity of Dissolved Oxygen (%)
A	7.8, 6.4, 8.2, 6.9
B	6.7, 6.8, 7.1, 6.9, 7.3
C	7.2, 7.4, 6.9, 6.4, 6.5
D	6.0, 7.4, 6.5, 6.9, 7.2, 6.8

Do the data indicate a significant difference in the average amount of dissolved oxygen for the four locations?

12.5 TWO-WAY ANALYSIS OF VARIANCE WITH ONE OBSERVATION PER CELL

In many practical problems one is interested in investigating the effects of two factors that influence an outcome. For example, the variety of grain and the type of fertilizer used both affect the yield of a plot; or the score on a standard examination is influenced by the size of the class and the instructor.

Let us suppose that two factors affect the outcome of an experiment. Suppose also that one observation is available at each of a number of levels of these two factors. Let $Y_{ij}(i = 1, 2, \dots, a;\ j = 1, 2, \dots, b)$ be the observation when the first factor is at the ith level and the second factor at the jth level. Assume that

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \qquad i = 1, 2, \ldots, a; \quad j = 1, 2, \ldots, b, \tag{1}$$

where α_i is the effect of the ith level of the first factor, β_j is the effect of the jth level of the second factor, and ε_{ij} is the random error, which is assumed to be normally distributed with mean 0 and variance σ^2. We will assume that the ε_{ij}'s are independent. It follows that Y_{ij} are independent normal RVs with means $\mu + \alpha_i + \beta_j$ and variance σ^2. There is no loss of generality in assuming that $\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = 0$, for if $\mu_{ij} = \mu' + \alpha_i' + \beta_j'$, we can write

$$\begin{aligned}\mu_{ij} &= (\mu' + \overline{\alpha}' + \overline{\beta}') + (\alpha_i' - \overline{\alpha}') + (\beta_j' - \overline{\beta}') \\ &= \mu + \alpha_i + \beta_j\end{aligned}$$

and $\sum_{i=1}^{a} \alpha_i = 0$, $\sum_{j=1}^{b} \beta_j = 0$. Here we have written $\overline{\alpha}'$ and $\overline{\beta}'$ for the means of α_i''s and β_j''s, respectively. Thus Y_{ij} may denote the yield from use of the ith variety of some grain and the jth type of some fertilizer. The two hypotheses of interest are

$$\alpha_1 = \alpha_2 = \cdots = \alpha_a = 0 \quad \text{and} \quad \beta_1 = \beta_2 = \cdots = \beta_b = 0.$$

The first of these, for example, says that the first factor has no effect on the outcome of the experiment.

In view of the fact that $\sum_{i=1}^{a} \alpha_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$, $\alpha_a = -\sum_{i=1}^{a-1} \alpha_i$, $\beta_b = -\sum_{j=1}^{b-1} \beta_j$, and we can write our model in matrix notation as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where

$$\begin{aligned}\mathbf{Y} &= (Y_{11}, Y_{12}, \ldots, Y_{1b}, Y_{21}, Y_{22}, \ldots, Y_{2b}, \ldots, Y_{a1}, Y_{a2}, \ldots, Y_{ab})', \\ \boldsymbol{\beta} &= (\mu, \alpha_1, \alpha_2, \ldots, \alpha_{a-1}, \beta_1, \beta_2, \ldots, \beta_{b-1})', \\ \boldsymbol{\varepsilon} &= (\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{1b}, \varepsilon_{21}, \varepsilon_{22}, \ldots, \varepsilon_{2b}, \ldots, \varepsilon_{a1}, \varepsilon_{a2}, \ldots, \varepsilon_{ab})',\end{aligned}$$

and

$$\mathbf{X} = \begin{array}{c} \begin{array}{c|cccc|cccc} \mu & \alpha_1 & \alpha_2 & \cdots & \alpha_{a-1} & \beta_1 & \beta_2 & \cdots & \beta_{b-1} \end{array} \\ \left(\begin{array}{c|cccc|cccc} 1 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & 0 & -1 & -1 & \cdots & -1 \\ 1 & 0 & 1 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 0 & -1 & -1 & \cdots & -1 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & -1 & -1 & \cdots & -1 & 1 & 0 & \cdots & 0 \\ 1 & -1 & -1 & \cdots & -1 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & -1 & -1 & \cdots & -1 & 0 & 0 & \cdots & 1 \\ 1 & -1 & -1 & \cdots & -1 & -1 & -1 & \cdots & -1 \end{array}\right) \end{array}.$$

The vector of unknown parameters $\boldsymbol{\beta}$ is $(a + b - 1) \times 1$, and the matrix $\mathbf{X}$ is $ab \times (a + b - 1)$ (b blocks of a rows each). We leave the reader to check that $\mathbf{X}$ is of full rank, $a + b - 1$. The hypothesis $H_\alpha\colon \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$ or $H_\beta\colon \beta_1 = \beta_2 = \cdots = \beta_b = 0$ can easily be put into the form $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$. For example, for H_β we can choose $\mathbf{H}$ to be the $(b - 1) \times (a + b - 1)$ matrix of full rank $b - 1$, given by

$$\mathbf{H} = \begin{array}{c} \begin{array}{c|cccc|cccc} \mu & \alpha_1 & \alpha_2 & \cdots & \alpha_{a-1} & \beta_1 & \beta_2 & \cdots & \beta_{b-1} \end{array} \\ \left(\begin{array}{c|cccc|cccc} 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{array}\right) \end{array}$$

Clearly, the model described above is a special case of the general linear hypothesis, and we can use Theorem 12.2.1 to test H_β.

To apply Theorem 12.2.1, we need the estimators $\hat{\mu}_{ij}$ and $\hat{\hat{\mu}}_{ij}$. It is easily checked that

$$\hat{\mu} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}}{ab} = \overline{y} \tag{3}$$

and

$$\hat{\alpha}_i = \overline{y}_{i\cdot} - \overline{y}, \qquad \hat{\beta}_j = \overline{y}_{\cdot j} - \overline{y}, \tag{4}$$

where $\overline{y}_{i\cdot} = \sum_{j=1}^{b} y_{ij}/b$, $\overline{y}_{\cdot j} = \sum_{i=1}^{a} y_{ij}/a$. Also, under H_β, for example,

$$\hat{\hat{\mu}} = \overline{y} \quad \text{and} \quad \hat{\hat{\alpha}}_i = \overline{y}_{i\cdot} - \overline{y}. \tag{5}$$

In the notation of Theorem 12.2.1, $n = ab$, $k = a + b - 1$, $r = b - 1$, so that $n - k = ab - a - b + 1 = (a-1)(b-1)$, and

$$F = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{i\cdot})^2 - \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y})^2}{\sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y})^2}. \tag{6}$$

Since

$$\begin{aligned} \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{i\cdot})^2 &= \sum_{i=1}^{a}\sum_{j=1}^{b}[(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y}) + (\overline{Y}_{\cdot j} - \overline{Y})]^2 \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y})^2 + a\sum_{j=1}^{b}(\overline{Y}_{\cdot j} - \overline{Y})^2, \end{aligned} \tag{7}$$

we may write

$$F = \frac{a\sum_{j=1}^{b}(\overline{Y}_{\cdot j} - \overline{Y})^2}{\sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y})^2}. \tag{8}$$

It follows that under H_β, $(a-1)F$ has a central $F(b-1, (a-1)(b-1))$ distribution.

The numerator of F in (8) measures the variability between the means $\overline{Y}_{\cdot j}$, and the denominator measures the variability that exists once the effects due to the two factors have been subtracted.

If H_α is the null hypothesis to be tested, one can show that under H_α the MLEs are

$$\hat{\hat{\mu}} = \overline{y} \quad \text{and} \quad \hat{\hat{\beta}}_j = \overline{y}_{\cdot j} - \overline{y}. \tag{9}$$

As before, $n = ab$, $k = a + b - 1$, but $r = a - 1$. Also,

$$(10)\quad F = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij}-\overline{Y}_{\cdot j})^2 - \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij}-\overline{Y}_{i\cdot}-\overline{Y}_{\cdot j}+\overline{Y})^2}{\sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij}-\overline{Y}_{i\cdot}-\overline{Y}_{\cdot j}+\overline{Y})^2},$$

which may be rewritten as

$$(11)\quad F = \frac{b\sum_{i=1}^{a}(\overline{Y}_{i\cdot}-\overline{Y})^2}{\sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij}-\overline{Y}_{i\cdot}-\overline{Y}_{\cdot j}+\overline{Y})^2}.$$

It follows that under H_α, $(b-1)F$ has a central $F(a-1,(a-1)(b-1))$ distribution. The numerator of F in (11) measures the variability between the means $\overline{Y}_{i\cdot}$.

If the data are put into the following form:

$\alpha \backslash \beta$		Level of factor 2: 1	2		b	Row mean
Level of factor 1	1	Y_{11},	Y_{12},	$\cdots$,	Y_{1b}	$\overline{Y}_{1\cdot}$
	2	Y_{21},	Y_{22},	$\cdots$,	Y_{2b}	$\overline{Y}_{2\cdot}$
	.	.	.	...	.	.
	.	.	.	...	.	.
	.	.	.	...	.	.
	a	Y_{a1},	Y_{a2},	$\cdots$,	Y_{ab}	$\overline{Y}_{a\cdot}$
Column mean		$\overline{Y}_{\cdot 1}$,	$\overline{Y}_{\cdot 2}$,	$\cdots$,	$\overline{Y}_{\cdot b}$	$\overline{Y}$

so that the rows represent various levels of factor 1, and the columns, the levels of factor 2, one can write

$$\begin{aligned}\text{between sum of squares for rows} &= b\sum_{i=1}^{a}(\overline{Y}_{i\cdot}-\overline{Y})^2\\ &= \text{sum of squares for factor 1}\\ &= \text{SS}_1.\end{aligned}$$

Similarly,

$$\begin{aligned}\text{between sum of squares for columns} &= a\sum_{j=1}^{b}(\overline{Y}_{\cdot j}-\overline{Y})^2\\ &= \text{sum of squares for factor 2}\\ &= \text{SS}_2.\end{aligned}$$

It is usual to write error or residual sum of squares (SSE) for the denominator of (8) or (11). These results are conveniently presented in an analysis of variance table as follows:

Two-Way Analysis of Variance Table with One Observation per Cell

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio
Rows	SS_1	$a-1$	$MS_1 = SS_1/(a-1)$	MS_1/MSE
Columns	SS_2	$b-1$	$MS_2 = SS_2/(b-1)$	MS_2/MSE
Error	SSE	$(a-1)(b-1)$	$MSE = SSE/(a-1)(b-1)$	
Mean	$ab\overline{Y}^2$	1	$ab\overline{Y}^2$	
Total	$\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2$	ab	$\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2/ab$	

Example 1. The following table gives the yield (pounds per plot) of three varieties of wheat, obtained with four different kinds of fertilizers.

	Variety of Wheat		
Fertilizer	A	B	C
α	8	3	7
β	10	4	8
γ	6	5	6
δ	8	4	7

Let us test the hypothesis of equality in the average yields of the three varieties of wheat and the null hypothesis that the four fertilizers are equally effective.

In our notation, $b = 3$, $a = 4$, $\overline{y}_1. = 6$, $\overline{y}_2. = 7.33$, $\overline{y}_3. = 5.67$, $\overline{y}_4. = 6.33$, $\overline{y}._1 = 8$, $\overline{y}._2 = 4$, $\overline{y}._3 = 7$, $\overline{y} = 6.33$.

Also,

$$\begin{aligned} SS_1 &= \text{sum of squares due to fertilizer} \\ &= 3[(.33)^2 + 1^2 + (0.66)^2 + 0^2] \\ &= 4.67; \\ SS_2 &= \text{sum of squares due to variety of wheat} \\ &= 4[(1.67)^2 + (2.33)^2 + (0.67)^2] \\ &= 34.67 \end{aligned}$$

and

$$\begin{aligned} SSE &= \sum_{i=1}^{4}\sum_{j=1}^{3}(y_{ij} - \overline{y}_i. - \overline{y}._j + \overline{y})^2 \\ &= 7.33 \end{aligned}$$

The results are shown in the following table:

Analysis of Variance

Source	SS	d.f.	MS	F-Ratio
Variety of wheat	34.67	2	17.33	14.2
Fertilizer	4.67	3	1.56	1.28
Error	7.33	6	1.22	
Mean	481.33	1	481.33	
Total	528.00	12	44.00	

Now $F_{2,6,0.05} = 5.14$ and $F_{3,6,0.05} = 4.76$. Since $14.2 > 5.14$, we reject H_β, that there is equality in the average yield of the three varieties; but since $1.28 \not> 4.76$, we accept H_α, that the four fertilizers are equally effective.

PROBLEMS 12.5

1. Show that the matrix $\mathbf{X}$ for the model defined in (2) is of full rank, $a + b - 1$.
2. Prove statements (3), (4), (5), and (9).
3. The following data represent the units of production per day turned out by four different brands of machines used by four machinists:

	Machinist			
Machine	A_1	A_2	A_3	A_4
B_1	15	14	19	18
B_2	17	12	20	16
B_3	16	18	16	17
B_4	16	16	15	15

Test whether the differences in the performances of the machinists are significant and also whether the differences in the performances of the four brands of machines are significant. Use $\alpha = 0.05$.

4. Students were classified into four ability groups, and three different teaching methods were employed. The following table gives the mean for four groups:

Ability Group	Teaching Method A	B	C
1	15	19	14
2	18	17	12
3	22	25	17
4	17	21	19

Test the hypothesis that the teaching methods yield the same results, that is, that the teaching methods are equally effective.

5. The following table shows the yield (pounds per plot) of four varieties of wheat obtained with three different kinds of fertilizers.

Fertilizer	Variety of Wheat A	B	C	D
α	8	3	6	7
β	10	4	5	8
γ	8	4	6	7

Test the hypotheses that the four varieties of wheat yield the same average yield and that the three fertilizers are equally effective.

12.6 TWO-WAY ANALYSIS OF VARIANCE WITH INTERACTION

The model described in Section 12.5 assumes that the two factors act independently, that is, are *additive*. In practice, this is an assumption that needs testing. In this section we allow for the possibility that the two factors might jointly affect the outcome; that is, there might be *interactions*. More precisely, if Y_{ij} is the observation in the (i, j)th cell, we will consider the model

$$Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij}, \tag{1}$$

where $\alpha_i (i = 1, 2, \dots, a)$ represent row effects (or effects due to factor 1), $\beta_j (j = 1, 2, \dots, b)$ represent column effects (or effects due to factor 2), and γ_{ij} represent interactions or joint effects. We assume that ε_{ij} are independently distributed as $\mathcal{N}(0, \sigma^2)$. We assume further that

$$\sum_{i=1}^{a} \alpha_i = 0 = \sum_{j=1}^{b} \beta_j \quad \text{and} \quad \begin{aligned} \sum_{j=1}^{b} \gamma_{ij} &= 0 \quad \text{for all } i, \\ \sum_{i=1}^{a} \gamma_{ij} &= 0 \quad \text{for all } j. \end{aligned} \tag{2}$$

The hypothesis of interest is

(3) $$H_0: \gamma_{ij} = 0 \qquad \text{for all } i, j.$$

One may also be interested in testing that all α's are 0 or that all β's are 0 in the presence of interactions γ_{ij}.

We first note that (2) is not restrictive since we can write

$$Y_{ij} = \mu' + \alpha_i' + \beta_j' + \gamma_{ij}' + \varepsilon_{ij},$$

where α_i', β_j', and γ_{ij}' do not satisfy (2), as

$$\begin{aligned} Y_{ij} &= \mu' + \overline{\alpha}' + \overline{\beta}' + \overline{\gamma}' + (\alpha_i' - \overline{\alpha}' + \overline{\gamma}_{i\cdot}' - \overline{\gamma}') + (\beta_j' - \overline{\beta}' + \overline{\gamma}_{\cdot j}' - \overline{\gamma}') \\ &\quad + (\gamma_{ij}' - \overline{\gamma}_{i\cdot}' - \overline{\gamma}_{\cdot j}' + \overline{\gamma}') + \varepsilon_{ij}, \end{aligned}$$

and then (2) is satisfied by choosing

$$\begin{aligned} \mu &= \mu' + \overline{\alpha}' + \overline{\beta}' + \overline{\gamma}', \\ \alpha_i &= \alpha_i' - \overline{\alpha}' + \overline{\gamma}_{i\cdot}' - \overline{\gamma}', \\ \beta_j &= \beta_j' - \overline{\beta}' + \overline{\gamma}_{\cdot j}' - \overline{\gamma}', \end{aligned}$$

and

$$\gamma_{ij} = \gamma_{ij}' - \overline{\gamma}_{i\cdot}' - \overline{\gamma}_{\cdot j}' + \overline{\gamma}'.$$

Here

$$\overline{\alpha}' = a^{-1}\sum_{i=1}^{a}\alpha_i', \quad \overline{\beta}' = b^{-1}\sum_{j=1}^{b}\beta_j', \quad \overline{\gamma}_{i\cdot}' = b^{-1}\sum_{j=1}^{b}\gamma_{ij}',$$

$$\overline{\gamma}_{\cdot j}' = a^{-1}\sum_{i=1}^{a}\gamma_{ij}', \quad \text{and} \quad \overline{\gamma}' = (ab)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\gamma_{ij}'.$$

Next note that unless we replicate, that is, take more than one observation per cell, there are no degrees of freedom left to estimate the error SS (see Remark 1).

Let Y_{ijs} be the sth observation when the first factor is at the ith level and the second factor at the jth level, $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, $s = 1, 2, \ldots, m (> 1)$. Then the model becomes as follows:

	Level of Factor 2			
Level of Factor 1	1	2	$\cdots$	b
1	y_{111}	y_{121}	$\cdots$	y_{1b1}
	.	.	$\cdots$	.
	.	.	$\cdots$	.
	.	.	$\cdots$	.
	y_{11m}	y_{12m}	$\cdots$	y_{1bm}
2	y_{211}	y_{221}	$\cdots$	y_{2b1}
	.	.	$\cdots$	.
	.	.	$\cdots$	.
	.	.	$\cdots$	.
	y_{21m}	y_{22m}	$\cdots$	y_{2bm}
	.	.	$\cdots$	.
	.	.	$\cdots$	.
	.	.	$\cdots$	.
a	y_{a11}	y_{a21}	$\cdots$	y_{ab1}
	.	.	$\cdots$	.
	.	.	$\cdots$	.
	.	.	$\cdots$	.
	y_{a1m}	y_{a2m}	$\cdots$	y_{abm}

$$Y_{ijs} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijs}, \tag{4}$$

$i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, and $s = 1, 2, \ldots, m$, where ε_{ijs}'s are independent $\mathcal{N}(0, \sigma^2)$. We assume that $\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} \gamma_{ij} = \sum_{j=1}^{b} \gamma_{ij} = 0$. Suppose that we wish to test $H_\alpha: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$. We leave the reader to check that model (4) is then a special case of the general linear hypothesis with $n = abm$, $k = ab$, $r = a - 1$, and $n - k = ab(m - 1)$.

Let us write

$$\overline{Y} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{s=1}^{m} Y_{ijs}}{n}, \quad \overline{Y}_{ij\cdot} = \frac{\sum_{s=1}^{m} Y_{ijs}}{m}, \tag{5}$$
$$\overline{Y}_{i\cdot\cdot} = \frac{\sum_{j=1}^{b} \sum_{s=1}^{m} Y_{ijs}}{mb}, \quad \overline{Y}_{\cdot j\cdot} = \frac{\sum_{i=1}^{a} \sum_{s=1}^{m} Y_{ijs}}{am}.$$

Then it can be easily checked that

$$\begin{cases} \hat{\mu} = \hat{\hat{\mu}} = \overline{Y}, \quad \hat{\hat{\alpha}}_i = \overline{Y}_{i\cdot\cdot} - \overline{Y}, \quad \hat{\beta}_j = \hat{\hat{\beta}}_j = \overline{Y}_{\cdot j\cdot} - \overline{Y}, \\ \hat{\gamma}_{ij} = \hat{\hat{\gamma}}_{ij} = \overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}. \end{cases} \tag{6}$$

It follows from Theorem 12.2.1 that

$$(7)\quad F = \frac{\sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot} + \overline{Y}_{i\cdot\cdot} - \overline{Y})^2 - \sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot})^2}{\sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot})^2}.$$

Since

$$\sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot} + \overline{Y}_{i\cdot\cdot} - \overline{Y})^2 = \sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot})^2 + \sum_i \sum_j \sum_s (\overline{Y}_{i\cdot\cdot} - \overline{Y})^2,$$

we can write (7) as

$$(8)\quad F = \frac{bm \sum_i (\overline{Y}_{i\cdot\cdot} - \overline{Y})^2}{\sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot})^2}.$$

Under H_α the statistic $[ab(m-1)/(a-1)]F$ has the central $F(a-1, ab(m-1))$ distribution, so that the likelihood ratio test rejects H_α if

$$(9)\quad \frac{ab(m-1)}{a-1} \frac{mb \sum_i (\overline{Y}_{i\cdot\cdot} - \overline{Y})^2}{\sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot})^2} > c.$$

A similar analysis holds for testing $H_\beta : \beta_1 = \beta_2 = \cdots = \beta_b$.

Next consider the test of hypothesis $H_\gamma : \gamma_{ij} = 0$ for all i, j, that is, that the two factors are independent and the effects are additive. In this case, $n = abm$, $k = ab$, $r = (a-1)(b-1)$, and $n - k = ab(m-1)$. It can be shown that

$$(10)\quad \hat{\hat{\mu}} = \overline{Y}, \quad \hat{\hat{\alpha}}_i = \overline{Y}_{i\cdot\cdot} - \overline{Y}, \quad \text{and} \quad \hat{\hat{\beta}}_j = \overline{Y}_{\cdot j\cdot} - \overline{Y}.$$

Thus

$$(11)\quad F = \frac{\sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y})^2 - \sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot})^2}{\sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot})^2}.$$

Now

$$\begin{aligned}
\sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y})^2
&= \sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot} + \overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y})^2 \\
&= \sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot})^2 + \sum_i \sum_j \sum_s (\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y})^2,
\end{aligned}$$

so that we may write

$$(12) \qquad F = \frac{\sum_i \sum_j \sum_s (\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y})^2}{\sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot})^2}.$$

Under H_γ, the statistic $\{(m-1)ab/[(a-1)(b-1)]\}F$ has the $F((a-1)(b-1), ab(m-1))$ distribution. The likelihood ratio test rejects H_γ if

$$(13) \qquad \frac{(m-1)ab}{(a-1)(b-1)} \frac{m \sum_i \sum_j (\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y})^2}{\sum_i \sum_j \sum_s (Y_{ijs} - \overline{Y}_{ij\cdot})^2} > c.$$

Let us write

$$\begin{aligned} SS_1 &= \text{sum of squares due to factor 1 (row sum of squares)} \\ &= bm \sum_{i=1}^{a} (\overline{Y}_{i\cdot\cdot} - \overline{Y})^2, \\ SS_2 &= \text{sum of squares due to factor 2 (column sum of squares)} \\ &= am \sum_{j=1}^{b} (\overline{Y}_{\cdot j\cdot} - \overline{Y})^2, \\ SSI &= \text{sum of squares due to interaction} \\ &= m \sum_{i=1}^{a} \sum_{j=1}^{b} (\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y})^2, \end{aligned}$$

and

$$\begin{aligned} SSE &= \text{sum of squares due to error (residual sum of squares)} \\ &= \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{s=1}^{m} (Y_{ijs} - \overline{Y}_{ij\cdot})^2. \end{aligned}$$

Then we may summarize the foregoing results in the following table.

Two-Way Analysis of Variance Table with Interaction

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio
Rows	SS_1	$a-1$	$MS_1 = SS_1/(a-1)$	MS_1/MSE
Columns	SS_2	$b-1$	$MS_2 = SS_2/(b-1)$	MS_2/MSE
Interaction	SSI	$(a-1)(b-1)$	$MSI = SSI/(a-1)(b-1)$	MSI/MSE
Error	SSE	$ab(m-1)$	$MSE = SSE/ab(m-1)$	
Mean	$abm\overline{X}^2$	1	$abm\overline{X}^2$	
Total	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{s=1}^{m} Y_{ijs}^2$	abm	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{s=1}^{m} Y_{ijs}^2/abm$	

Remark 1. Note that if $m = 1$, there are no d.f.'s associated with the SSE. Indeed, SSE $= 0$ if $m = 1$. Hence we cannot make tests of hypotheses when $m = 1$, and for this reason we assume that $m > 1$.

Example 1. To test the effectiveness of three different teaching methods, three instructors were randomly assigned 12 students each. The students were then randomly assigned to the different teaching methods and were taught exactly the same material. At the conclusion of the experiment, identical examinations were given to the students with the following results in regard to grades:

Teaching Method	Instructor I	Instructor II	Instructor III
1	95	60	86
	85	90	77
	74	80	75
	74	70	70
2	90	89	83
	80	90	70
	92	91	75
	82	86	72
3	70	68	74
	80	73	86
	85	78	91
	85	93	89

From the data the table of means is as follows:

	$\overline{y}_{ij\cdot}$			$\overline{y}_{i\cdot\cdot}$
	82	75	77	78.0
	86	89	75	83.3
	80	78	85	81.0
$\overline{y}_{\cdot j\cdot}$	82.7	80.7	79.0	$\overline{y} = 80.8$

Then

$$\begin{aligned} SS_1 &= \text{sum of squares due to methods} \\ &= bm\sum_{i=1}^{a}(\overline{y}_{i\cdot\cdot} - \overline{y})^2 \\ &= 3 \times 4 \times 14.13 = 169.56, \end{aligned}$$

$$\begin{aligned}
SS_2 &= \text{sum of squares due to instructors} \\
&= am \sum_{j=1}^{b} (\overline{y}_{\cdot j \cdot} - \overline{y})^2 \\
&= 3 \times 4 \times 6.86 = 82.32, \\
SSI &= \text{sum of squares due to interaction} \\
&= m \sum_{i=1}^{3} \sum_{j=1}^{3} (\overline{y}_{ij\cdot} - \overline{y}_{i\cdot\cdot} - \overline{y}_{\cdot j \cdot} + \overline{y})^2 \\
&= 4 \times 140.45 = 561.80,
\end{aligned}$$

and

$$\begin{aligned}
SSE &= \text{residual sum of squares} \\
&= \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{s=1}^{4} (y_{ijs} - \overline{y}_{ij\cdot})^2 = 1830.00.
\end{aligned}$$

Analysis of Variance

Source	SS	d.f.	MSS	F-Ratio
Methods	169.56	2	84.78	1.25
Instructors	82.32	2	41.16	0.61
Interactions	561.80	4	140.45	2.07
Error	1830.00	27	67.78	

With $\alpha = 0.05$, we see from the tables that $F_{2,27,0.05} = 3.35$ and $F_{4,27,0.05} = 2.73$, so that we cannot reject any of the three hypotheses that the three methods are equally effective, that the three instructors are equally effective, and that the interactions are all 0.

PROBLEMS 12.6

1. Prove statement (6).
2. Obtain the likelihood ratio test of the null hypothesis H_β: $\beta_1 = \beta_2 = \cdots = \beta_b = 0$.
3. Prove statement (10).
4. Suppose that the following data represent the units of production turned out each day by three different machinists, each working on the same machine for three different days:

	Machinist		
Machine	*A*	*B*	*C*
B_1	15, 15, 17	19, 19, 16	16, 18, 21
B_2	17, 17, 17	15, 15, 15	19, 22, 22
B_3	15, 17, 16	18, 17, 16	18, 18, 18
B_4	18, 20, 22	15, 16, 17	17, 17, 17

Using a 0.05 level of significance, test whether (a) the differences among the machinists are significant, (b) the differences among the machines are significant, and (c) the interactions are significant.

5. In an experiment to determine whether four different makes of automobiles average the same gasoline mileage, a random sample of two cars of each make was taken from each of four cities. Each car was then test run on 5 gallons of gasoline of the same brand. The following table gives the number of miles traveled.

	Automobile Make			
City	*A*	*B*	*C*	*D*
Cleveland	92.3, 104.1	90.4, 103.8	110.2, 115.0	120.0, 125.4
Detroit	96.2, 98.6	91.8, 100.4	112.3, 111.7	124.1, 121.1
San Francisco	90.8, 96.2	90.3, 89.1	107.2, 103.8	118.4, 115.6
Denver	98.5, 97.3	96.8, 98.8	115.2, 110.2	126.2, 120.4

Construct the analysis of variance table. Test the hypothesis of no automobile effect, no city effect, and no interactions. Use $\alpha = 0.05$.