Chapter 9
Analysis of Covariance

Traditionally, analysis of covariance (ACOVA) has been used as a tool in the analysis of designed experiments. Suppose one or more measurements are made on a group of experimental units. In an agricultural experiment, such a measurement might be the amount of nitrogen in each plot of ground prior to the application of any treatments. In animal husbandry, the measurements might be the height and weight of animals before treatments are applied. One way to use such information is to create blocks of experimental units that have similar values of the measurements. Analysis of covariance uses a different approach. In analysis of covariance, an experimental design is chosen that does not depend on these supplemental observations. The concomitant observations come into play as regression variables that are added to the basic experimental design model.

The goal of analysis of covariance is the same as the goal of blocking. The regression variables are used to *reduce the variability of treatment comparisons*. In this traditional context, comparisons among treatments remain the primary goal of the analysis. Exercises 9.1 and 9.5 are important practical illustrations of how this is accomplished. Snedecor and Cochran (1980, Chapter 18) discuss the practical uses of analysis of covariance. Cox (1958, Chapter 4) discusses the proper role of concomitant observations in experimental design. *Biometrics* has devoted two entire issues to analysis of covariance: Volume 13, Number 3, 1957 and Volume 38, Number 3, 1982.

From a theoretical point of view, analysis of covariance involves the analysis of a model with a partitioned model matrix, say

$$Y = [X, Z]\begin{bmatrix}\beta\\ \gamma\end{bmatrix} + e, \tag{1}$$

where X is an $n \times p$ matrix, Z is an $n \times s$ matrix, $\mathrm{E}(e) = 0$, and $\mathrm{Cov}(e) = \sigma^2 I$. Analysis of covariance is a technique for analyzing model (1) based on the analysis of the reduced model

$$Y = X\delta + e. \tag{2}$$

R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer Texts in Statistics, DOI 10.1007/978-1-4419-9816-3_9,

The point is that model (2) should be a model whose analysis is relatively easy. In traditional applications, X is taken as the model matrix for a balanced analysis of variance. The Z matrix can be anything, but traditionally consists of columns of regression variables.

The practical application of general linear model theory is prohibitively difficult without a computer program to perform the worst of the calculations. There are, however, special cases: notably, simple linear regression, one-way ANOVA, and balanced multifactor ANOVA, in which the calculations are not prohibitive. Analysis of covariance allows computations of the BLUEs and the *SSE* for model (1) by performing several analyses on tractable special cases plus finding the generalized inverse of an $s \times s$ matrix. Since finding the generalized inverse of anything bigger than, say, a 3×3 matrix is difficult for hand calculations, one would typically not want more than three columns in the Z matrix for such purposes.

As mentioned earlier, in the traditional application of performing an ANOVA while adjusting for the effect of some regression variables, the primary interest is in the ANOVA. The regression variables are there only to sharpen the analysis. The inference on the ANOVA part of the model is performed after fitting the regression variables. To test whether the regression variables really help to sharpen the analysis, they should be tested after fitting the ANOVA portion of the model. The basic computation for performing these tests is finding the *SSE* for model (1). This implicitly provides a method for finding the *SSE* for submodels of model (1). Appropriate tests are performed by comparing the *SSE* for model (1) to the *SSE*s of the various submodels.

Sections 1 and 2 present the theory of estimation and testing for general *partitioned models*. Sections 3 and 4 present nontraditional applications of the theory. Section 3 applies the analysis of covariance results to the problem of fixing up balanced ANOVA problems that have lost their balance due to the existence of some missing data. Although applying analysis of covariance to missing data problems is not a traditional experimental design application, it is an application that was used for quite some time until computational improvements made it largely unnecessary. Section 4 uses the analysis of covariance results to derive the analysis for balanced incomplete block designs. Section 5 presents Milliken and Graybill's (1970) test of a linear model versus a nonlinear alternative.

I might also add that I personally find the techniques of Sections 1 and 2 to be some of the most valuable tools available for deriving results in linear model theory.

9.1 Estimation of Fixed Effects

To obtain least squares estimates, we break the estimation space of model (9.0.1) into two orthogonal parts. As usual, let M be the perpendicular projection operator onto $C(X)$. Note that $C(X,Z) = C[X,(I-M)Z]$. One way to see this is that from model (9.0.1)

$$\mathrm{E}(Y) = X\beta + Z\gamma = X\beta + MZ\gamma + (I-M)Z\gamma = [X, MZ]\begin{bmatrix}\beta\\ \gamma\end{bmatrix} + (I-M)Z\gamma.$$

Since $C(X) = C([X, MZ])$, clearly model (9.0.1) holds if and only if $\mathrm{E}(Y) \in C[X, (I-M)Z]$.

Let P denote the perpendicular projection matrix onto $C([X,Z]) = C([X,(I-M)Z])$. Since the two sets of column vectors in $[X,(I-M)Z]$ are orthogonal, the perpendicular projection matrix for the entire space is the sum of the perpendicular projection matrices for the subspaces $C(X)$ and $C[(I-M)Z]$, cf. Theorem B.45. Thus,

$$P = M + (I-M)Z\left[Z'(I-M)Z\right]^{-} Z'(I-M)$$

and write

$$M_{(I-M)Z} \equiv (I-M)Z\left[Z'(I-M)Z\right]^{-} Z'(I-M).$$

Least squares estimates satisfy $X\hat{\beta} + Z\hat{\gamma} = PY = MY + M_{(I-M)Z}Y$.

We now consider estimation of estimable functions of γ and β. The formulae are simpler if we incorporate the estimate

$$\hat{\gamma} = \left[Z'(I-M)Z\right]^{-} Z'(I-M)Y, \tag{1}$$

which we will later show to be a least squares estimate.

First consider an estimable function of γ, say, $\xi'\gamma$. For this to be estimable, there exists a vector ρ such that $\xi'\gamma = \rho'[X\beta + Z\gamma]$. For this equality to hold for all β and γ, we must have $\rho'X = 0$ and $\rho'Z = \xi'$. The least squares estimate of $\xi'\gamma$ is

$$\begin{aligned}\rho'PY &= \rho'\left\{M + M_{(I-M)Z}\right\}Y\\ &= \rho'MY + \rho'(I-M)Z\left[Z'(I-M)Z\right]^{-} Z'(I-M)Y\\ &= 0 + \rho'Z\hat{\gamma}\\ &= \xi'\hat{\gamma}.\end{aligned}$$

The penultimate equality stems from the fact that $\rho'X = 0$ implies $\rho'M = 0$.

An arbitrary estimable function of β, say, $\lambda'\beta$, has $\lambda'\beta = \rho'[X\beta + Z\gamma]$ for some ρ. For this equality to hold for all β and γ, we must have $\rho'X = \lambda'$ and $\rho'Z = 0$. As a result, the least squares estimate is

$$\begin{aligned}\rho'PY &= \rho'\left\{M + M_{(I-M)Z}\right\}Y\\ &= \rho'MY + \rho'(I-M)Z\left[Z'(I-M)Z\right]^{-} Z'(I-M)Y\\ &= \rho'MY + \rho'(I-M)Z\hat{\gamma}\\ &= \rho'MY - \rho'MZ\hat{\gamma}\\ &= \rho'M(Y - Z\hat{\gamma})\\ &\equiv \lambda'\hat{\beta}.\end{aligned}$$

Define

$$X\hat{\beta} = M(Y - Z\hat{\gamma}). \tag{2}$$

We now establish that

$$X\hat{\beta} + Z\hat{\gamma} = PY$$

so that $\hat{\gamma}$ is a least squares estimate of γ and and $X\hat{\beta}$ is a least squares estimate of $X\beta$. Write

$$\begin{aligned} X\hat{\beta} + Z\hat{\gamma} &= M(Y - Z\hat{\gamma}) + Z\hat{\gamma} \\ &= MY + (I - M)Z\hat{\gamma} \\ &= MY + M_{(I-M)Z}Y = PY. \end{aligned}$$

Often in ACOVA, the X matrix comes from a model that is simple to analyze, like one-way ANOVA or a balanced multifactor ANOVA. If the model $Y = X\beta + e$ has simple formula for computing an estimate of some function $\lambda'\beta = \rho'X\beta$, say a contrast, then that simple formula must be incorporated into $\rho'MY$. Under conditions that we will explore, $\lambda'\beta$ is also an estimable function under the ACOVA model $Y = X\beta + Z\gamma + e$ and, with the same vector ρ, the estimate is $\lambda'\hat{\beta} = \rho'M(Y - Z\hat{\gamma})$. That means that the same (simple) computational procedure that was applied to the data Y in order to estimate $\lambda'\beta$ in $Y = X\beta + e$ can also be applied to $Y - Z\hat{\gamma}$ to estimate $\lambda'\beta$ in $Y = X\beta + Z\gamma + e$, see Exercise 9.1. We now explore the conditions necessary to make this happen, along with other issues related to estimability in ACOVA models.

Often Z consists of columns of regression variables, in which case $Z'(I-M)Z$ is typically nonsingular. In that case, both γ and $X\beta$ are estimable. In particular,

$$\gamma = \left[Z'(I-M)Z\right]^{-1} Z'(I-M)[X\beta + Z\gamma]$$

with estimate

$$\begin{aligned} \hat{\gamma} &= \left[Z'(I-M)Z\right]^{-1} Z'(I-M)PY \\ &= \left[Z'(I-M)Z\right]^{-1} Z'(I-M)[M + M_{(I-M)Z}]Y \\ &= \left[Z'(I-M)Z\right]^{-1} Z'(I-M)M_{(I-M)Z}Y \\ &= \left[Z'(I-M)Z\right]^{-1} Z'(I-M)Y. \end{aligned}$$

X is traditionally the model matrix for an ANOVA model, so β is usually not estimable. However, when $Z'(I-M)Z$ is nonsingular, $X\beta$ is estimable in the ACOVA model. Observe that

$$\begin{aligned} &\left\{I - Z\left[Z'(I-M)Z\right]^{-1} Z'(I-M)\right\}[X\beta + Z\gamma] \\ &= X\beta + Z\gamma - Z\left[Z'(I-M)Z\right]^{-1} Z'(I-M)Z\gamma \\ &= X\beta + Z\gamma - Z\gamma \\ &= X\beta. \end{aligned}$$

Thus, in the nonsingular case, anything that is estimable in $Y = X\beta + e$ is also estimable in the ACOVA model. In particular, if $\lambda' = \rho' X$, the estimate of $\lambda'\beta$ in $Y = X\beta + e$ is $\rho' MY$. Clearly, for the ACOVA model,

$$\rho'\left\{I - Z\left[Z'(I-M)Z\right]^{-1}Z'(I-M)\right\}[X\beta + Z\gamma] = \rho' X\beta = \lambda'\beta$$

and the estimate is

$$\lambda'\hat{\beta} = \rho'\left\{I - Z\left[Z'(I-M)Z\right]^{-1}Z'(I-M)\right\}PY.$$

We now show that $\lambda'\hat{\beta} = \rho' M(Y - Z\hat{\gamma})$. As mentioned earlier, the beauty of this result is that if we know how to estimate $\lambda'\beta$ in $Y = X\beta + e$ using Y, exactly the same method applied to $Y - Z\hat{\gamma}$ will give the estimate in the ACOVA model.

As discussed earlier, an estimable function $\lambda'\beta$, has $\lambda'\beta = \tilde{\rho}'[X\beta + Z\gamma]$ for some $\tilde{\rho}$ with $\tilde{\rho}'X = \lambda'$ and $\tilde{\rho}'Z = 0$. Also as before, the least squares estimate is

$$\tilde{\rho}'PY = \tilde{\rho}'M(Y - Z\hat{\gamma}).$$

In the nonsingular case, if ρ is any vector that has $\rho' X = \lambda'$, we can turn it into a vector $\tilde{\rho}$ that has both $\tilde{\rho}'X = \lambda'$ and $\tilde{\rho}'Z = 0$, simply by defining

$$\tilde{\rho}' = \rho'\left\{I - Z\left[Z'(I-M)Z\right]^{-1}Z'(I-M)\right\}.$$

Moreover,

$$\tilde{\rho}'M(Y - Z\hat{\gamma}) = \rho' M(Y - Z\hat{\gamma}),$$

so the same estimation procedure applied to Y in $Y = X\beta + e$ gets applied to $Y - Z\hat{\gamma}$ in $Y = X\beta + Z\gamma + e$ when estimating the estimable function $\lambda'\beta$.

In general, if $[Z'(I-M)Z]$ is singular, neither γ nor $X\beta$ are estimable. The estimable functions of γ will be those that are linear functions of $(I-M)Z\gamma$. This is shown below.

Proposition 9.1.1. $\xi'\gamma$ is estimable if and only if $\xi' = \rho'(I-M)Z$ for some vector ρ.

PROOF. If $\xi'\gamma$ is estimable, there exists ρ such that $\xi'\gamma = \rho'[X, Z]\begin{bmatrix}\beta\\ \gamma\end{bmatrix}$, so $\rho'[X, Z] = (0, \xi')$ and $\rho' X = 0$. Therefore, $\xi' = \rho' Z = \rho'(I-M)Z$. Conversely, if $\xi' = \rho'(I-M)Z$ then $\rho'(I-M)[X, Z]\begin{bmatrix}\beta\\ \gamma\end{bmatrix} = \xi'\gamma$. □

Proposition 9.1.1 is phrased in terms of estimating a function of γ, but it also applies with appropriate changes to estimation of β.

Finally, if $X\beta$ and γ are estimable, that is, if $(I-M)Z$ is of full rank, it is easy to see that

$$\mathrm{Cov}\begin{bmatrix} X\hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \sigma^2 \begin{bmatrix} M + MZ\left[Z'(I-M)Z\right]^{-1} Z'M & -MZ\left[Z'(I-M)Z\right]^{-1} \\ -\left[Z'(I-M)Z\right]^{-1} Z'M & \left[Z'(I-M)Z\right]^{-1} \end{bmatrix}.$$

9.2 Estimation of Error and Tests of Hypotheses

The estimate of the variance σ^2 is the MSE. We will find $SSE = Y'(I-P)Y$ in terms of M and Z. The error sum of squares is

$$\begin{aligned} Y'(I-P)Y &= Y'\Big[(I-M) - (I-M)Z\left[Z'(I-M)Z\right]^{-} Z'(I-M)\Big]Y \qquad (1) \\ &= Y'(I-M)Y - Y'(I-M)Z\left[Z'(I-M)Z\right]^{-} Z'(I-M)Y. \end{aligned}$$

Using the notation $E_{AB} = A'(I-M)B$, we have

$$Y'(I-P)Y = E_{YY} - E_{YZ}E_{ZZ}^{-}E_{ZY}.$$

EXAMPLE 9.2.1. Consider a balanced two-way analysis of variance with no replication and one covariate (regression variable, concomitant variable, supplemental observation). The analysis of covariance model can be written

$$y_{ij} = \mu + \alpha_i + \eta_j + \gamma z_{ij} + e_{ij},$$

$i = 1,\ldots,a,\ j = 1,\ldots,b$. Thus X is the model matrix for the balanced two-way ANOVA without replication,

$$y_{ij} = \mu + \alpha_i + \eta_j + e_{ij},$$

$i = 1,\ldots,a,\ j = 1,\ldots,b$, and Z is an $ab \times 1$ matrix that contains the values of z_{ij}. The sum of squares for error in the covariate analysis is $E_{YY} - E_{YZ}^2/E_{ZZ}$, where

$$E_{YY} = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}\right)^2,$$

$$E_{YZ} = E_{ZY} = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}\right)\left(z_{ij} - \bar{z}_{i\cdot} - \bar{z}_{\cdot j} + \bar{z}_{\cdot\cdot}\right),$$

$$E_{ZZ} = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(z_{ij} - \bar{z}_{i\cdot} - \bar{z}_{\cdot j} + \bar{z}_{\cdot\cdot}\right)^2.$$

Tests for analysis of covariance models are found by considering the reductions in sums of squares for error due to the models. For instance, if $C(X_0) \subset C(X)$ and we want to test the reduced model

$$Y = X_0\beta_0 + Z\gamma + e$$

against the full model (9.0.1), the test statistic is

$$\frac{[Y'(I-P_0)Y - Y'(I-P)Y]\,/\,[r(X,Z) - r(X_0,Z)]}{[Y'(I-P)Y]\,/\,[n - r(X,Z)]},$$

where P_0 is the perpendicular projection operator onto $C(X_0,Z)$. We have already found $Y'(I-P)Y$. If M_0 is the perpendicular projection operator onto $C(X_0)$,

$$Y'(I-P_0)Y = Y'(I-M_0)Y - Y'(I-M_0)Z\left[Z'(I-M_0)Z\right]^{-}Z'(I-M_0)Y.$$

Often these computations can be facilitated by writing an analysis of covariance table.

EXAMPLE 9.2.1 CONTINUED. The analysis of covariance table is given below in matrix notation. Recall that, for example,

$$\begin{aligned} Y'M_\alpha Y &= b\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2, \\ Y'M_\alpha Z &= b\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot}), \\ Z'M_\alpha Z &= b\sum_{i=1}^{a}(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})^2. \end{aligned}$$

ACOVA Table

Source	df	SS_{YY}	SS_{YZ}	SS_{ZZ}
Grand Mean	1	$Y'\frac{1}{n}J_n^nY$	$Y'\frac{1}{n}J_n^nZ$	$Z'\frac{1}{n}J_n^nZ$
Treatments (α)	$a-1$	$Y'M_\alpha Y$	$Y'M_\alpha Z$	$Z'M_\alpha Z$
Treatments (η)	$b-1$	$Y'M_\eta Y$	$Y'M_\eta Z$	$Z'M_\eta Z$
Error	$n-a-b+1$	$Y'(I-M)Y$	$Y'(I-M)Z$	$Z'(I-M)Z$

If we want to test $H_0: \eta_1 = \eta_2 = \cdots = \eta_b$, the error sum of squares under the reduced model is

$$\begin{aligned} &\left[Y'(I-M)Y + Y'M_\eta Y\right] - \left[Y'(I-M)Z + Y'M_\eta Z\right] \\ &\qquad \times \left[Z'(I-M)Z + Z'M_\eta Z\right]^{-}\left[Z'(I-M)Y + Z'M_\eta Y\right]. \end{aligned}$$

All of these terms are available from the ACOVA table. With more than one covariate, the terms in the SS_{YZ} and SS_{ZZ} columns would be matrices and it would be more involved to compute $[Z'(I-M)Z + Z'M_\eta Z]^{-}$.

Exercise 9.1 Consider a one-way ANOVA with one covariate. The model is

$$y_{ij} = \mu + \alpha_i + \xi x_{ij} + e_{ij},$$

$i = 1, \ldots, t$, $j = 1, \ldots, N_i$. Find the BLUE of the contrast $\sum_{i=1}^{t} \lambda_i \alpha_i$. Find the variance of the contrast.

Exercise 9.2 Consider the problem of estimating β_p in the regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i. \tag{2}$$

Let r_i be the *ordinary residual* from fitting

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \cdots + \alpha_{p-1} x_{ip-1} + e_i$$

and s_i be the residual from fitting

$$x_{ip} = \gamma_0 + \gamma_1 x_{i1} + \cdots + \gamma_{p-1} x_{ip-1} + e_i.$$

Show that the least squares estimate of β_p is $\hat{\xi}$ from fitting the model

$$r_i = \xi s_i + e_i, \quad i = 1, \ldots, n, \tag{3}$$

that the *SSE* from models (2) and (3) are the same, and that $(\hat{\beta}_0, \ldots, \hat{\beta}_{p-1})' = \hat{\alpha} - \hat{\beta}_p \hat{\gamma}$ with $\hat{\alpha} = (\hat{\alpha}_0, \ldots, \hat{\alpha}_{p-1})'$ and $\hat{\gamma} = (\hat{\gamma}_0, \ldots, \hat{\gamma}_{p-1})'$. Discuss the usefulness of these results for computing regression estimates. (These are the key results behind the *sweep operator* that is often used in regression computations.) What happens to the results if r_i is replaced by y_i in model (3)?

Exercise 9.3 Suppose $\lambda_1'\beta$ and $\lambda_2'\gamma$ are estimable in model (9.0.1). Use the normal equations to find find least squares estimates of $\lambda_1'\beta$ and $\lambda_2'\gamma$.

Hint: Reparameterize the model as $X\beta + Z\gamma = X\delta + (I - M)Z\gamma$ and use the normal equations on the reparameterized model. Note that $X\delta = X\beta + MZ\gamma$.

Exercise 9.4 Derive the test for model (9.0.1) versus the reduced model $Y = X\beta + Z_0\gamma_0 + e$, where $C(Z_0) \subset C(Z)$. Describe how the procedure would work for testing $H_0 : \gamma_2 = 0$ in the model $y_{ij} = \mu + \alpha_i + \eta_j + \gamma_1 z_{ij1} + \gamma_2 z_{ij2} + e_{ij}$, $i = 1, \ldots, a$, $j = 1, \ldots, b$.

Exercise 9.5 An experiment was conducted with two treatments. There were four levels of the first treatment and five levels of the second treatment. Besides the data y, two covariates were measured, x_1 and x_2. The data are given below. Analyze the data with the assumption that there is no interaction between the treatments.

i	j	y_{ij}	x_{1ij}	x_{2ij}	i	j	y_{ij}	x_{1ij}	x_{2ij}
1	1	27.8	5.3	9	3	1	22.4	3.0	13
	2	27.8	5.2	11		2	21.0	4.5	12
	3	26.2	3.6	13		3	30.6	5.4	18
	4	24.8	5.2	17		4	25.4	6.6	21
	5	17.8	3.6	10		5	15.9	4.1	9
2	1	19.6	4.7	12	4	1	14.1	5.4	10
	2	28.4	5.8	17		2	29.5	6.8	18
	3	26.3	3.3	22		3	29.2	5.3	22
	4	18.3	4.1	8		4	21.5	6.2	9
	5	20.8	5.7	11		5	25.5	6.4	22

9.3 Application: Missing Data

When a few observations are missing from, say, a balanced multifactor design, the balance is lost and the analysis would seem to be quite complicated. One use of the analysis of covariance is to allow the analysis with missing data to be performed using results for the original balanced design.

Consider an original design

$$Y = X\beta + e,$$

with $Y = (y_1, \ldots, y_n)'$. For each missing observation y_i, include a covariate $z_i = (0, \ldots, 0, 1, 0, \ldots, 0)'$ with the 1 in the ith place. Set each y_i that is missing equal to zero.

We wish to show that the *SSE* in this ACOVA model equals the *SSE* in the model with the missing observations deleted. The *MSE* in the covariance model will also equal the *MSE* in the model with deletions. In the covariance model, although we are artificially adding observations by setting missing observations to zero, we are also removing those degrees of freedom from the error by adding covariates.

Suppose r observations are missing. Without loss of generality, we can assume that the last r observations are missing. The $n \times r$ matrix of covariates can be written

$$Z = \begin{bmatrix} 0 \\ I_r \end{bmatrix},$$

where I_r is an $r \times r$ identity matrix and 0 is an $(n-r) \times r$ matrix of zeros. Let X be the $n \times p$ model matrix for the model with no missing observations and let X_* be the $(n-r) \times p$ model matrix for the model with the missing observations deleted. Again we can assume that

$$X = \begin{bmatrix} X_* \\ X_r \end{bmatrix},$$

where X_r is the $r \times p$ matrix whose rows are the rows of X corresponding to the missing observations. The analysis of covariance model

$$Y = X\beta + Z\gamma + e$$

can now be written as

$$Y = \begin{bmatrix} X_* & 0 \\ X_r & I_r \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + e.$$

Notice that

$$C\left(\begin{bmatrix} X_* & 0 \\ X_r & I_r \end{bmatrix}\right) = C\left(\begin{bmatrix} X_* & 0 \\ 0 & I_r \end{bmatrix}\right).$$

Let M_* be the perpendicular projection operator onto $C(X_*)$ and let P be the perpendicular projection operator onto

$$C\left(\begin{bmatrix} X_* & 0 \\ X_r & I_r \end{bmatrix}\right).$$

It is easy to see that

$$P = \begin{bmatrix} M_* & 0 \\ 0 & I_r \end{bmatrix}.$$

Writing Y as $Y' = [Y_*', 0]$, we find that

$$Y'(I-P)Y = \begin{bmatrix} Y_*' & 0 \end{bmatrix} \begin{bmatrix} I - M_* & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Y_* \\ 0 \end{bmatrix} = Y_*'(I - M_*)Y_*.$$

Since $Y'(I-P)Y$ is the SSE from the covariate model and $Y_*'(I-M_*)Y_*$ is the SSE for the model with the missing observations dropped, we are done. Note that the values we put in for the missing observations do not matter for computing the SSE. Tests of hypotheses can be conducted by comparing SSEs for different models.

With $Y' = [Y_*', 0]$, estimation will be the same in both models. The least squares estimate of

$$\begin{bmatrix} X_*\beta \\ X_r\beta + \gamma \end{bmatrix}$$

is

$$PY = \begin{bmatrix} M_*Y_* \\ 0 \end{bmatrix}.$$

Any estimable function in the model $Y_* = X_*\beta + e_*$ is estimable in the covariate model, and the estimates are the same. The function $\rho_*'X_*\beta$ equals $\rho'(X\beta + Z\gamma)$, where $\rho' = [\rho_*', 0]$ and 0 is a $1 \times r$ matrix of zeros. Thus, $\rho'PY = \rho_*'M_*Y_*$.

Exercise 9.6 Show that

$$P = \begin{bmatrix} M_* & 0 \\ 0 & I_r \end{bmatrix}.$$

An alternative approach to the missing value problem is based on finding substitutes for the missing values. The substitutes are chosen so that if one acts like the substitutes are real data, the correct *SSE* is computed. (The degrees of freedom for error must be corrected.)

Setting the problem up as before, we have

$$Y = X\beta + Z\gamma + e,$$

and the estimate $\hat{\gamma}$ can be found. It is proposed to treat $Y - Z\hat{\gamma} = [Y_*', -\hat{\gamma}']'$ as the data from an experiment with model matrix X. Recalling that $\hat{\gamma} = [Z'(I-M)Z]^{-1}Z'(I-M)Y$, we see that the *SSE* from this analysis is

$$\begin{aligned}
&(Y - Z\hat{\gamma})'(I-M)(Y - Z\hat{\gamma}) \\
&= Y'(I-M)Y - 2\hat{\gamma}'Z'(I-M)Y + \hat{\gamma}'Z'(I-M)Z\hat{\gamma} \\
&= Y'(I-M)Y - 2Y'(I-M)Z[Z'(I-M)Z]^{-1}Z'(I-M)Y \\
&\quad + Y'(I-M)Z[Z'(I-M)Z]^{-1}[Z'(I-M)Z][Z'(I-M)Z]^{-1}Z'(I-M)Y \\
&= Y'(I-M)Y - Y'(I-M)Z[Z'(I-M)Z]^{-1}Z'(I-M)Y,
\end{aligned}$$

which is the *SSE* from the covariate analysis. Therefore, if we replace missing values by $-\hat{\gamma}$ and do the regular analysis, we get the correct *SSE*.

This procedure also gives the same estimates as the ACOVA procedure. Using this method, the estimate of $X\beta$ is $M(Y - Z\hat{\gamma})$ just as in (9.1.2) of the ACOVA procedure. The variance of an estimate, say $\rho'X\hat{\beta}$, needs to be calculated as in an analysis of covariance. It is $\sigma^2(\rho'M\rho + \rho'MZ[Z'(I-M)Z]^{-1}Z'M\rho)$, not the naive value of $\sigma^2\rho'M\rho$.

For $r = 1$ missing observation and a variety of balanced designs, formulae have been obtained for $-\hat{\gamma}$ and are available in many statistical methods books.

Exercise 9.7 Derive $-\hat{\gamma}$ for a randomized complete block design when $r = 1$.

9.4 Application: Balanced Incomplete Block Designs

The analysis of covariance technique can be used to develop the analysis of a balanced incomplete block design. Suppose that a design is to be set up with b blocks and t treatments, but the number of treatments that can be observed in any block is k, where $k < t$. One natural way to proceed would be to find a design where each pair of treatments occurs together in the same block a fixed number of times, say λ. Such a design is called a *balanced incomplete block (BIB) design*.

Let r be the common number of replications for each treatment. There are two well-known facts about the parameters introduced so far. First, the total number of experimental units in the design must be the number of blocks times the number of units in each block, i.e., bk, but the total number of units must also be the number of treatments times the number of times we observe each treatment, i.e., tr; thus

$$tr = bk. \tag{1}$$

Second, the number of within block comparisons between any given treatment and the other treatments is fixed. One way to count this is to multiply the number of other treatments $(t-1)$ by the number of times each occurs in a block with the given treatment (λ). Another way to count it is to multiply the number of other treatments in a block $(k-1)$ times the number of blocks that contain the given treatment (r). Therefore,

$$(t-1)\lambda = r(k-1). \tag{2}$$

EXAMPLE 9.4.1. An experiment was conducted to examine the effects of fertilizers on potato yields. Six treatments (A, B, C, D, E, and F) were used but blocks were chosen that contained only five experimental units. The experiment was performed using a balanced incomplete block design with six blocks. The potato yields (in pounds) along with the mean yield for each block are reported in Table 9.1.

Table 9.1 Potato Yields in Pounds for Six Fertilizer Treatments.

Block	Data					Block Means
1	E 583	B 512	F 661	A 399	C 525	536.0
2	B 439	C 460	D 424	E 497	F 592	482.4
3	A 334	E 466	C 492	B 431	D 355	415.6
4	F 570	D 433	E 514	C 448	A 344	461.8
5	D 402	A 417	B 420	F 626	E 615	496.0
6	C 450	F 490	A 268	D 375	B 347	386.0

The six treatments consist of all of the possible combinations of two factors. One factor was that a nitrogen-based fertilizer was either applied (n_1) or not applied (n_0). The other factor was that a phosphate-based fertilizer was either not applied (p_0), applied in a single dose (p_1), or applied in a double dose (p_2). In terms of the factorial structure, the six treatments are $A = n_0p_0$, $B = n_0p_1$, $C = n_0p_2$, $D = n_1p_0$, $E = n_1p_1$, and $F = n_1p_2$. From the information in Table 9.1, it is a simple matter to check that $t = 6$, $b = 6$, $k = 5$, $r = 5$, and $\lambda = 4$. After deriving the theory for balanced incomplete block designs, we will return to these data and analyze them.

The balanced incomplete block model can be written as

$$y_{ij} = \mu + \beta_i + \tau_j + e_{ij}, \qquad e_{ij}\text{s i.i.d. } N(0,\sigma^2), \tag{3}$$

where $i = 1,\ldots,b$ and $j \in D_i$ or $j = 1,\ldots,t$ and $i \in A_j$. D_i is the set of indices for the treatments in block i. A_j is the set of indices for the blocks in which treatment j occurs. Note that there are k elements in each set D_i and r elements in each set A_j.

In applying the analysis of covariance, we will use the balanced one-way ANOVA determined by the grand mean and the blocks to help analyze the model with covariates. The covariates are taken as the columns of the model matrix associated with the treatments. Writing (3) in matrix terms, we get

$$Y = [X,Z]\begin{bmatrix}\beta\\ \tau\end{bmatrix} + e,$$

$$\beta' = (\mu, \beta_1, \ldots, \beta_b), \qquad \tau' = (\tau_1, \ldots, \tau_t).$$

Note that in performing the analysis of covariance for this model our primary interest lies in the coefficients of the covariates, i.e., the treatment effects. To perform an analysis of covariance, we need to find $Y'(I-M)Z$ and $[Z'(I-M)Z]^-$.

First, find $Z'(I-M)Z$. There are t columns in Z; write $Z = [Z_1, \ldots, Z_t]$. The rows of the mth column indicate the presence or absence of the mth treatment. Using two subscripts to denote the rows of vectors, we can write the mth column as

$$Z_m = [z_{ij,m}], \qquad \text{where } z_{ij,m} = \delta_{jm}$$

and δ_{jm} is 1 if $j = m$, and 0 otherwise. In other words, Z_m is 0 for all rows except the r rows that correspond to an observation on treatment m; those r rows are 1.

To get the $t \times t$ matrix $Z'(I-M)Z$, we find each individual element of $Z'Z$ and $Z'MZ$. This is done by finding $Z_s'Z_m$ and $Z_s'M'MZ_m$ for all values of m and s. If $m = s$, we get

$$Z_m'Z_m = \sum_i \sum_j (z_{ij,m})^2 = \sum_{j=1}^t \sum_{i \in A_j} \delta_{jm} = \sum_{j=1}^t r\delta_{jm} = r.$$

Now, if $m \neq s$, because each observation has only one treatment associated with it, either $z_{ij,s}$ or $z_{ij,m}$ equals 0; so for $s \neq m$,

$$Z_s'Z_m = \sum_i \sum_j (z_{ij,s})(z_{ij,m}) = \sum_{j=1}^t \sum_{i \in A_j} \delta_{js}\delta_{jm} = \sum_{j=1}^t r\delta_{js}\delta_{jm} = 0.$$

Thus the matrix $Z'Z$ is rI_t, where I_t is a $t \times t$ identity matrix.

Recall that in this problem, X is the model matrix for a one-way ANOVA where the "treatments" of the ANOVA are the blocks of the BIB design and there are k observations on each "treatment." Using two subscripts to denote each row and each column of a matrix, we can write the projection matrix as in Section 4.1,

$$M = [v_{ij,i'j'}], \qquad \text{where } v_{ij,i'j'} = \frac{1}{k}\delta_{ii'}.$$

Let

$$MZ_m = [d_{ij,m}].$$

Then

$$
\begin{aligned}
d_{ij,m} &= \sum_{i'j'} v_{ij,i'j'} z_{i'j',m} \\
&= \sum_{j'=1}^{t} \sum_{i' \in A_{j'}} \frac{1}{k} \delta_{ii'} \delta_{j'm} \\
&= \sum_{j'=1}^{t} \delta_{j'm} \sum_{i' \in A_{j'}} \frac{1}{k} \delta_{ii'} \\
&= \sum_{i' \in A_m} \frac{1}{k} \delta_{ii'} \\
&= \frac{1}{k} \delta_i(A_m),
\end{aligned}
$$

where $\delta_i(A_m)$ is 1 if $i \in A_m$ and 0 otherwise. In other words, if treatment m is in block i, then all k of the units in block i have $d_{ij,m} = 1/k$. If treatment m is not in block i, all k of the units in block i have $d_{ij,m} = 0$. Since treatment m is contained in exactly r blocks,

$$
\begin{aligned}
Z'_m M' M Z_m &= \sum_{ij} (d_{ij,m})^2 = \sum_{i=1}^{b} \sum_{j \in D_i} k^{-2} \delta_i(A_m) \\
&= \sum_{i=1}^{b} (k/k^2) \delta_i(A_m) = \frac{r}{k}.
\end{aligned}
$$

Since, for $s \neq m$, there are λ blocks in which both treatments s and m are contained,

$$
\begin{aligned}
Z'_s M' M Z_m &= \sum_{ij} (d_{ij,s})(d_{ij,m}) = \sum_{i=1}^{b} \sum_{j \in D_i} (1/k^2) \delta_i(A_s) \delta_i(A_m) \\
&= \sum_{i=1}^{b} (k/k^2) \delta_i(A_s) \delta_i(A_m) = \frac{\lambda}{k}.
\end{aligned}
$$

It follows that the matrix $Z'MZ$ has values r/k down the diagonal and values λ/k off the diagonal. This can be written as

$$
Z'MZ = \frac{1}{k} \left[(r-\lambda) I + \lambda J_t^t \right].
$$

Finally, we can now write

$$
\begin{aligned}
Z'(I-M)Z &= Z'Z - Z'MZ \\
&= rI - k^{-1} \left[(r-\lambda) I + \lambda J_t^t \right] \\
&= k^{-1} \left[(r(k-1) + \lambda) I - \lambda J_t^t \right].
\end{aligned}
$$

This matrix can be simplified further. Define

$$W = I - (1/t)J_t^t. \tag{4}$$

Note that W is a perpendicular projection operator and that equation (2) gives

$$r(k-1) + \lambda = \lambda t.$$

With these substitutions, we obtain

$$Z'(I-M)Z = (\lambda/k)\left[tI - J_t^t\right] = (\lambda t/k)\left[I - (1/t)J_t^t\right] = (\lambda t/k)W.$$

We need to find a generalized inverse of $Z'(I-M)Z$. Because W is a projection operator, it is easily seen that

$$\left[Z'(I-M)Z\right]^- = (k/\lambda t)W. \tag{5}$$

We also need to be able to find the $1 \times t$ vector $Y'(I-M)Z$. The vector $(I-M)Y$ has elements $(y_{ij} - \bar{y}_{i\cdot})$, so

$$Y'(I-M)Z_m = \sum_{ij}(y_{ij} - \bar{y}_{i\cdot})z_{ij,m} = \sum_{j=1}^{t}\delta_{jm}\sum_{i\in A_j}(y_{ij} - \bar{y}_{i\cdot}) = \sum_{i\in A_m}(y_{im} - \bar{y}_{i\cdot}).$$

Define

$$Q_m = \sum_{i\in A_m}(y_{im} - \bar{y}_{i\cdot}).$$

Then

$$Y'(I-M)Z = (Q_1, \ldots, Q_t).$$

Since the β effects are for blocks, our primary interests are in estimable functions $\xi'\tau$ and in estimating σ^2. From (9.1.1) and Proposition 9.1.1, write

$$\xi' = \rho'(I-M)Z$$

to get

$$\xi'\hat{\tau} = \rho'(I-M)Z\left[Z'(I-M)Z\right]^- Z'(I-M)Y$$

and, from (9.2.1),

$$SSE = Y'(I-M)Y - Y'(I-M)Z\left[Z'(I-M)Z\right]^- Z'(I-M)Y.$$

Both of these formulae involve the term $(I-M)Z\left[Z'(I-M)Z\right]^-$. This term can be simplified considerably. Note that since the columns of Z are 0s and 1s, indicating the presence or absence of a treatment effect,

$$ZJ_t^1 = J_n^1.$$

Because M is defined from a one-way ANOVA,

$$0 = (I-M)J_n^1 = (I-M)ZJ_t^1. \tag{6}$$

From (5), (4), and (6), it is easily seen that

$$(I-M)Z\left[Z'(I-M)Z\right]^{-} = (k/\lambda t)(I-M)Z.$$

Using this fact, we get that the BLUE of $\xi'\tau$ is

$$\begin{aligned}
\xi'\hat{\tau} &= \rho'(I-M)Z\left[Z'(I-M)Z\right]^{-}Z'(I-M)Y \\
&= \rho'(k/\lambda t)(I-M)ZZ'(I-M)Y \\
&= (k/\lambda t)\xi'(Q_1,\ldots,Q_t)' \\
&= (k/\lambda t)\sum_{j=1}^{t}\xi_j Q_j.
\end{aligned}$$

Many computer programs, e.g., Minitab, present *adjusted treatment means* $(k/\lambda t)Q_j + \bar{y}_{\cdot\cdot}$ that can be used to estimate contrasts, because $\sum_j \xi_j \bar{y}_{\cdot\cdot} = 0$. The variance of the estimate of the contrast is

$$\begin{aligned}
\mathrm{Var}\left(\xi'\hat{\tau}\right) &= \sigma^2\rho'(I-M)Z\left[Z'(I-M)Z\right]^{-}Z'(I-M)\rho \\
&= \sigma^2(k/\lambda t)\xi'\xi.
\end{aligned}$$

From the estimate and the variance, it is a simple matter to see that

$$SS\left(\xi'\tau\right) = (k/\lambda t)\left[\sum_{j=1}^{t}\xi_j Q_j\right]^2 \Big/ \xi'\xi.$$

The error sum of squares is

$$\begin{aligned}
SSE &= Y'(I-M)Y - Y'(I-M)Z\left[Z'(I-M)Z\right]^{-}Z'(I-M)Y \\
&= Y'(I-M)Y - (k/\lambda t)Y'(I-M)ZZ'(I-M)Y \\
&= \sum_{ij}(y_{ij}-\bar{y}_{i\cdot})^2 - \frac{k}{\lambda t}\sum_{j=1}^{t}Q_j^2.
\end{aligned}$$

Exercise 9.8 Show that $\xi'\tau$ is estimable if and only if $\xi'\tau$ is a contrast.

Hint: One direction is easy. For the other direction, show that for $\xi' = (\xi_1,\ldots,\xi_t)$,

$$\xi' = (k/\lambda t)\xi'Z'(I-M)Z.$$

Exercise 9.9 Show that if $\xi'\tau$ and $\eta'\tau$ are contrasts and that if $\xi'\eta = 0$, then $\xi'\tau = 0$ and $\eta'\tau = 0$ put orthogonal constraints on $C(X,Z)$, i.e., the treatment sum of squares can be broken down with orthogonal contrasts in the usual way.

Hint: Let $\xi' = \rho_1'[X,Z]$ and $\eta' = \rho_2'[X,Z]$. Show that

$$\rho_1'(I-M)Z\left[Z'(I-M)Z\right]^- Z'(I-M)\rho_2 = 0.$$

Suppose that the treatments have quantitative levels, say $x_1, \ldots, x_t$, that are equally spaced. Model (3) can be reparameterized as

$$y_{ij} = \mu + \beta_i + \gamma_1 x_j + \gamma_2 x_j^2 + \cdots + \gamma_{t-1} x_j^{t-1} + e_{ij}.$$

We would like to show that the orthogonal polynomial contrasts for the balanced incomplete block design are the same as for a balanced one-way ANOVA. In other words, tabled polynomial contrasts, which are useful in balanced ANOVAs, can also be used to analyze balanced incomplete block designs. More generally, the treatments in a BIB may have a factorial structure with quantitative levels in some factor (e.g., Example 9.4.1). We would like to establish that the polynomial contrasts in the factor can be used in the usual way to analyze the data.

Because this is a balanced incomplete block design, Z is the model matrix for a balanced one-way ANOVA (without a grand mean). As in Section 7.3, define orthogonal polynomials $T = ZB$ by ignoring blocks. (Here we are not interested in J as an orthogonal polynomial, so we take B as a $t \times t-1$ matrix.) Write $B = [b_1, \ldots, b_{t-1}]$. If treatments are levels of a single quantitative factor, then the b_js are tabled orthogonal polynomial contrasts. If the treatments have factorial structure, the b_js are obtained from tabled contrasts as in the continuation of Example 9.4.1 below. The important fact is that the b_js are readily obtained. Note that $J_t'b_j = 0$ for all j, and $b_i'b_j = 0$ for $i \neq j$.

A model with treatments replaced by regression variables can be written $Y = X\beta + T\eta + e$, where $\eta = (\eta_1, \ldots, \eta_{t-1})'$. For a simple treatment structure, η_j would be the coefficient for a jth degree polynomial. For a factorial treatment structure, η_j could be the coefficient for some cross-product term. The key points are that the hypothesis $\eta_j = 0$ corresponds to some hypothesis that can be interpreted as in Section 6.7 or Section 7.3, and that the columns of T are orthogonal.

As we have seen, the model $Y = X\beta + T\eta + e$ is equivalent to the model $Y = X\delta + (I-M)T\eta + e$, where η is identical in the two models. Thus the test of $\eta_j = 0$ can be performed in the second model. In the second model, $\hat{\eta}$ is independent of $\hat{\delta}$ because of the orthogonality. If the columns of $(I-M)T$ are orthogonal, the estimates of the η_js are independent. Finally, and most importantly, we can show that the contrast in the τs that corresponds to testing $\eta_j = 0$ is simply $b_j'\tau$, where $\tau = (\tau_1, \ldots, \tau_t)'$.

To show that the columns of $(I-M)T$ are orthogonal, it suffices to show that $T'(I-M)T$ is diagonal.

$$\begin{aligned} T'(I-M)T &= B'Z'(I-M)ZB \\ &= (\lambda t/k)B'WB \\ &= (\lambda t/k)B'B. \end{aligned}$$

The last equality follows from the definition of W and the fact that $J_t'b_j = 0$ for all j. Note that $B'B$ is diagonal because $b_i'b_j = 0$ for $i \neq j$.

Finally, the contrast that corresponds to testing $\eta_j = 0$ is $\rho'(I-M)Z\tau$, where ρ is the jth column of $(I-M)T$, i.e., $\rho = (I-M)Zb_j$. This is true because $(I-M)T$ has orthogonal columns. The contrast is then

$$\begin{aligned}[(I-M)Zb_j]'(I-M)Z\tau &= b_j'Z'(I-M)Z\tau \\ &= (\lambda t/k)b_j'W\tau \\ &= (\lambda t/k)b_j'\tau\end{aligned}$$

or, equivalently, the contrast is

$$b_j'\tau.$$

We now apply these results to the analysis of the data in Example 9.4.1.

EXAMPLE 9.4.1 CONTINUED. The computation of the Q_ms is facilitated by the following table.

Treatment	n_0p_0	n_0p_1	n_0p_2	n_1p_0	n_1p_1	n_1p_2
$\sum_{i\in A_m} y_{im}$	1762.0	2149.0	2375.0	1989.0	2675.0	2939.0
$\sum_{i\in A_m} \bar{y}_{i\cdot}$	2295.4	2316.0	2281.8	2241.8	2391.8	2362.2
Q_m	−533.4	−167.0	93.2	−252.8	283.2	576.8

An analysis of variance table can be computed.

ANOVA

Source	df	SS	MS	F
Blocks (Ignoring Trts)	5	74857.77	14971.553	
Treatments (After Blks)	5	166228.98	33245.797	31.97
Error	19	19758.22	1039.906	
Total	29	260844.97		

Clearly, there are significant differences among the treatments. These can be explored further by examining contrasts. The factorial structure of the treatments suggests looking at nitrogen effects, phosphate effects, and interaction effects. With two levels of nitrogen, the only available contrast is $(1)n_0 + (-1)n_1$. Phosphate was applied at quantitative levels 0, 1, and 2. The linear contrast in phosphate is $(-1)p_0 + (0)p_1 + (1)p_2$. The quadratic contrast in phosphate is $(1)p_0 + (-2)p_1 + (1)p_2$. Combining these to obtain interaction contrasts and rewriting them as contrasts in the original six treatments gives $b_1, \dots, b_5$.

b_js

Treatments	n_0p_0	n_0p_1	n_0p_2	n_1p_0	n_1p_1	n_1p_2
N	1	1	1	−1	−1	−1
P linear	−1	0	1	−1	0	1
P quadratic	1	−2	1	1	−2	1
$N-P$ linear	−1	0	1	1	0	−1
$N-P$ quadratic	1	−2	1	−1	2	−1

Source	df	SS	F
N	1	51207.20	49.24
P linear	1	110443.67	106.21
P quadratic	1	2109.76	2.03
$N-P$ linear	1	2146.30	2.06
$N-P$ quadratic	1	322.06	0.31

The conclusions to be drawn are clear. There is a substantial increase in yields due to adding the nitrogen-based fertilizer. For the dosages of phosphate used, there is a definite increasing linear relationship between amount of phosphate and yield of potatoes. There is no evidence of any interaction.

Note that the linear relationship between phosphate and yield is an approximation that holds in some neighborhood of the dosages used in the experiment. It is well known that too much fertilizer will actually kill most plants. In particular, no potato plant will survive having an entire truckload of phosphate dumped on it.

Exercise 9.10 Derive the analysis for a Latin square with one row missing.
Hint: This problem is at the end of Section 9.4, not Section 9.3.

Exercise 9.11 Eighty wheat plants were grown in each of 5 different fields. Each of 6 individuals (A, B, C, D, E, and F) were asked to pick 8 "representative" plants in each field and measure the plants's heights. Measurements were taken on 6 different days. The data consist of the differences between the mean height of the 8 "representative" plants and the mean of all the heights in the field on that day. Thus the data measure the bias in selecting "representative" plants. The exact design and the data are given below. Analyze the data. (Although they are barely recognizable as such, these data are from Cochran and Cox, 1957.)

	Field				
Day	1	2	3	4	5
1	E 3.50	A 0.75	C 2.28	F 1.77	D 2.28
2	D 3.78	B 1.46	A −1.06	E 1.46	F 2.76
3	F 2.32	C 2.99	B −0.28	D 1.18	E 3.39
4	C 4.13	D 4.02	E 1.81	B 1.46	A 1.50
5	A 1.38	E 1.65	F 2.64	C 2.60	B 1.50
6	B 1.22	F 2.83	D 1.57	A −1.30	C 1.97

9.5 Application: Testing a Nonlinear Full Model

Consider testing the model

$$Y = X\beta + e, \qquad e \sim N(0, \sigma^2 I), \tag{1}$$

against a nonlinear model that involves a matrix function of $X\beta$, say

$$Y = X\beta + Z(X\beta)\gamma + e.$$

We assume that the matrix $Z(X\beta)$ has constant rank. More precisely, we need to assume that $r[(I-M)Z(X\beta)]$ and $r[X, Z(X\beta)]$ are constant functions of β. If each column of $Z(v)$ is a distinct nonlinear function of v, these conditions often hold.

Milliken and Graybill (1970) developed an exact F test for this problem. A similar result appears in the first (1965) edition of Rao (1973). When $Z(v) \equiv [v_1^2, \ldots, v_n^2]'$, the procedure gives Tukey's famous one degree of freedom for nonadditivity test, cf. Tukey (1949) or Christensen (1996a, Section 10.4). Tests for Mandel's (1961, 1971) extensions of the Tukey model also fit in this class of tests. Christensen and Utts (1992) extended these tests to log-linear and logit models, and Christensen (1997, Section 7.3) examines the Tukey and Mandel models in the context of log-linear models. St. Laurent (1990) showed that Milliken and Graybill's test is equivalent to a score test and thus shares the asymptotic properties of the generalized likelihood ratio test. St. Laurent also provides references to other applications of this class of tests.

To develop the test, fit model (1) to get $\tilde{Y} \equiv MY$ and define

$$\tilde{Z} \equiv Z(\tilde{Y}).$$

Now fit the model

$$Y = X\beta + \tilde{Z}\gamma + e$$

treating $\tilde{Z}$ as a known model matrix that does not depend on Y. Let P be the perpendicular projection operator onto $C(X, \tilde{Z})$, so the usual F test for $H_0 : \gamma = 0$ is based on

$$F = \frac{Y'M_{(I-M)\tilde{Z}}Y/r[(I-M)\tilde{Z}]}{Y'(I-P)Y/[n-r(X,\tilde{Z})]} \sim F(r[(I-M)\tilde{Z}], n-r[X,\tilde{Z}]). \tag{2}$$

To show that (2) really holds under the null hypothesis of model (1), consider the distribution of Y given $\tilde{Y}$. Write

$$Y = MY + (I-M)Y = \tilde{Y} + (I-M)Y.$$

Under the null model, $\tilde{Y}$ and $(I-M)Y$ are independent and

$$(I-M)Y \sim N[0, \sigma^2(I-M)],$$

so

$$Y|\tilde{Y} \sim N[\tilde{Y}, \sigma^2(I-M)].$$

Use the general results from Section 1.3 that involve checking conditions like $VAVAV = VAV$ and $VAVBV = 0$ to establish that the F statistic has the stated $F(r[(I-M)\tilde{Z}], n-r[X,\tilde{Z}])$ distribution conditional on $\tilde{Y}$. Finally, by assumption, the degrees of freedom for the F distribution do not depend on $\tilde{Y}$, so the conditional distribution does not depend on $\tilde{Y}$ and it must also be the unconditional distribution.

Exercise 9.12 Prove that display (2) is true.

9.6 Additional Exercises

Exercise 9.6.1 Sulzberger (1953) and Williams (1959) examined the maximum compressive strength parallel to the grain (y) of 10 hoop trees and how it was affected by temperature. A covariate, the moisture content of the wood (x), was also measured. Analyze the data, which are reported below.

Temperature in Celsius

	−20		0		20		40		60	
Tree	y	x	y	x	y	x	y	x	y	x
1	13.14	42.1	12.46	41.1	9.43	43.1	7.63	41.4	6.34	39.1
2	15.90	41.0	14.11	39.4	11.30	40.3	9.56	38.6	7.27	36.7
3	13.39	41.1	12.32	40.2	9.65	40.6	7.90	41.7	6.41	39.7
4	15.51	41.0	13.68	39.8	10.33	40.4	8.27	39.8	7.06	39.3
5	15.53	41.0	13.16	41.2	10.29	39.7	8.67	39.0	6.68	39.0
6	15.26	42.0	13.64	40.0	10.35	40.3	8.67	40.9	6.62	41.2
7	15.06	40.4	13.25	39.0	10.56	34.9	8.10	40.1	6.15	41.4
8	15.21	39.3	13.54	38.8	10.46	37.5	8.30	40.6	6.09	41.8
9	16.90	39.2	15.23	38.5	11.94	38.5	9.34	39.4	6.26	41.7
10	15.45	37.7	14.06	35.7	10.74	36.7	7.75	38.9	6.29	38.2

Exercise 9.6.2 Suppose that in Exercise 7.7.1 on motor oil pricing, the observation on store 7, brand H was lost. Treat the stores as blocks in a randomized complete block design. Plug in an estimate of the missing value and analyze the data without correcting the $MSTrts$ or any variance estimates. Compare the results of this approximate analysis to the results of a correct analysis.

Exercise 9.6.3 The missing value procedure that consists of analyzing the model $(Y - Z\hat{\gamma}) = X\beta + e$ has been shown to give the correct SSE and BLUEs; however, sums of squares explained by the model are biased upwards. For a randomized complete block design with a treatments and b blocks and the observation in the c,d cell missing, show that the correct mean square for treatments is the naive (biased) mean square treatments minus $[y_{\cdot d} - (a-1)\hat{y}_{cd}]^2/a(a-1)^2$, where $y_{\cdot d}$ is the sum of all actual observations in block d, and $\hat{y}_{cd}$ is the pseudo-observation (the nonzero element of $Z\hat{\gamma}$).

Exercise 9.6.4 State whether each design given below is a balanced incomplete block design, and if so, give the values of b, t, k, r, and λ.

(a) The experiment involves 5 treatments: A, B, C, D, and E. The experiment is laid out as follows.

Block	Treatments	Block	Treatments
1	A,B,C	6	A,B,D
2	A,B,E	7	A,C,D
3	A,D,E	8	A,C,E
4	B,C,D	9	B,C,E
5	C,D,E	10	B,D,E

(b) The following design has 9 treatments: A, B, C, D, E, F, G, H, and I.

Block	Treatments	Block	Treatments
1	B,C,D,G	6	C,D,E,I
2	A,C,E,H	7	A,D,H,I
3	A,B,F,I	8	B,E,G,I
4	A,E,F,G	9	C,F,G,H
5	B,D,F,H		

(c) The following design has 7 treatments: A, B, C, D, E, F, G.

Block	Treatments	Block	Treatments
1	C,E,F,G	5	B,C,D,G
2	A,D,F,G	6	A,C,D,E
3	A,B,E,G	7	B,D,E,F
4	A,B,C,F		