hw7 Linear Model

Zhuodiao Kuang

2025-10

1 1

4.24. Suppose Y_1, Y_2, and Y_3 are measurements of the angles of a triangle subject to error. The information is given as a linear model $Y_i = \theta_i + \varepsilon_i$, where θ_i are the true angles, $i = 1, 2, 3$. Assume that $\mathrm{E}(\varepsilon_i) = 0$, and $\mathrm{Var}(\varepsilon_i) = \sigma^2$. Obtain the least squares estimates of θ_i (subject to the constraint $\theta_1 + \theta_2 + \theta_3 = 180°$).

The least squares estimates for the true angles θ_i are given by:

$$\hat{\theta}_i = Y_i - \frac{1}{3}\left(\sum_{j=1}^{3} Y_j - 180°\right) \quad \text{for } i = 1, 2, 3$$

This result means each estimated angle ($\hat{\theta}_i$) is the original measurement (Y_i) adjusted by one-third of the total error from the required sum of 180°.

1.0.1 Derivation

To find the least squares estimates under a constraint, we use the method of **Lagrange multipliers**. The goal is to minimize the sum of squared errors between the measurements and the true values, subject to the geometric constraint that the angles of a triangle must sum to 180°.

1. Define the Objective Function

The sum of squared errors (SSE), which we want to minimize, is the function S:

$$S(\theta_1, \theta_2, \theta_3) = \sum_{i=1}^{3} (Y_i - \theta_i)^2$$

2. State the Constraint

The true angles must sum to 180 degrees:

$$\theta_1 + \theta_2 + \theta_3 = 180°$$

We can write this constraint as a function set to zero: $g(\theta_1, \theta_2, \theta_3) = \theta_1 + \theta_2 + \theta_3 - 180 = 0$.

3. Form the Lagrangian Function

The Lagrangian function, $\mathcal{L}$, combines the objective function and the constraint:

$$\mathcal{L}(\theta_1, \theta_2, \theta_3, \lambda) = S(\theta_1, \theta_2, \theta_3) + \lambda \cdot g(\theta_1, \theta_2, \theta_3)$$
$$\mathcal{L} = (Y_1 - \theta_1)^2 + (Y_2 - \theta_2)^2 + (Y_3 - \theta_3)^2 + \lambda(\theta_1 + \theta_2 + \theta_3 - 180)$$

4. Find the Partial Derivatives

To find the minimum, we take the partial derivative of $\mathcal{L}$ with respect to each θ_i and λ and set them to zero.

- $\frac{\partial \mathcal{L}}{\partial \theta_1} = -2(Y_1 - \theta_1) + \lambda = 0 \implies Y_1 - \theta_1 = \frac{\lambda}{2}$
- $\frac{\partial \mathcal{L}}{\partial \theta_2} = -2(Y_2 - \theta_2) + \lambda = 0 \implies Y_2 - \theta_2 = \frac{\lambda}{2}$
- $\frac{\partial \mathcal{L}}{\partial \theta_3} = -2(Y_3 - \theta_3) + \lambda = 0 \implies Y_3 - \theta_3 = \frac{\lambda}{2}$
- $\frac{\partial \mathcal{L}}{\partial \lambda} = \theta_1 + \theta_2 + \theta_3 - 180 = 0$

5. Solve for the Estimates

From the first three equations, we can express each $\hat{\theta}_i$ in terms of Y_i and λ:

$$\hat{\theta}_i = Y_i - \frac{\lambda}{2}$$

Now, substitute these expressions into the fourth equation (the constraint):

$$\left(Y_1 - \frac{\lambda}{2}\right) + \left(Y_2 - \frac{\lambda}{2}\right) + \left(Y_3 - \frac{\lambda}{2}\right) = 180$$
$$(Y_1 + Y_2 + Y_3) - 3\frac{\lambda}{2} = 180$$

Solving for $\frac{\lambda}{2}$:

$$3\frac{\lambda}{2} = (Y_1 + Y_2 + Y_3) - 180$$
$$\frac{\lambda}{2} = \frac{1}{3}(Y_1 + Y_2 + Y_3 - 180)$$

Finally, substitute this expression for $\frac{\lambda}{2}$ back into the equation for $\hat{\theta}_i$ to get the final estimates:

$$\hat{\theta}_i = Y_i - \frac{1}{3}\left(\sum_{j=1}^{3} Y_j - 180^\circ\right)$$

2 2

4.26. Consider the model $\mathbf{y} = \boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{V})$, where $\mathbf{V}$ is an $N \times N$ p.d. matrix. If $\mathbf{z} = \mathbf{B}\mathbf{y}$, and $\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\varepsilon}$, where $\mathbf{B} = \mathbf{V}^{-1/2}$, show that the linear model can be expressed as $\mathbf{z} = \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\eta}$, with $\boldsymbol{\eta} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_N)$. Obtain the least squares estimate of $\boldsymbol{\beta}$. Also obtain the estimate of $\boldsymbol{\beta}$ under the constraint $\mathbf{C}'\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{C}$ is an $N \times q$ matrix of rank $q < N$.

This problem involves transforming a generalized least squares (GLS) problem into an ordinary least squares (OLS) problem and then finding both the unconstrained and constrained estimators for the parameter vector $\boldsymbol{\beta}$.

2.0.1 1. Model Transformation

First, we show that the original model can be expressed in the specified transformed form.

1. **Original Model:** The model is given by $\mathbf{y} = \boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with the error term $\boldsymbol{\varepsilon}$ following a multivariate normal distribution $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{V})$.

2. **Apply the Transformation:** We are given the transformation matrix $\mathbf{B} = \mathbf{V}^{-1/2}$. Let's left-multiply the original model equation by $\mathbf{B}$:
$$\mathbf{B}\mathbf{y} = \mathbf{B}(\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$
$$\mathbf{B}\mathbf{y} = \mathbf{B}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\varepsilon}$$

3. **Substitute New Variables:** Define $\mathbf{z} = \mathbf{B}\mathbf{y}$ and $\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\varepsilon}$. Substituting these into the equation gives the transformed linear model:
$$\mathbf{z} = \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\eta}$$

4. **Find the Distribution of the New Error Term $\boldsymbol{\eta}$:** We need to find the mean and covariance of $\boldsymbol{\eta}$.

   - **Mean:** The expected value of $\boldsymbol{\eta}$ is:
   $$E[\boldsymbol{\eta}] = E[\mathbf{B}\boldsymbol{\varepsilon}] = \mathbf{B}E[\boldsymbol{\varepsilon}] = \mathbf{B}\mathbf{0} = \mathbf{0}$$

   - **Covariance:** The covariance matrix of $\boldsymbol{\eta}$ is:
   $$\text{Cov}(\boldsymbol{\eta}) = \text{Cov}(\mathbf{B}\boldsymbol{\varepsilon}) = \mathbf{B}\text{Cov}(\boldsymbol{\varepsilon})\mathbf{B}'$$
   Since $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$ and $\mathbf{B} = \mathbf{V}^{-1/2}$, we have:
   $$\text{Cov}(\boldsymbol{\eta}) = \mathbf{V}^{-1/2}(\sigma^2\mathbf{V})(\mathbf{V}^{-1/2})'$$
   Because $\mathbf{V}$ is a positive definite (and thus symmetric) matrix, its square root $\mathbf{V}^{1/2}$ and inverse square root $\mathbf{V}^{-1/2}$ are also symmetric. Therefore, $(\mathbf{V}^{-1/2})' = \mathbf{V}^{-1/2}$.
   $$\text{Cov}(\boldsymbol{\eta}) = \sigma^2(\mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2}) = \sigma^2\mathbf{V}^{(-1/2+1-1/2)} = \sigma^2\mathbf{V}^0 = \sigma^2\mathbf{I}_N$$
   Since $\boldsymbol{\eta}$ is a linear transformation of the normally distributed vector $\boldsymbol{\varepsilon}$, $\boldsymbol{\eta}$ is also normally distributed. Thus, we have shown that $\boldsymbol{\eta} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_N)$.

2.0.2 2. Unconstrained Least Squares Estimate

The transformed model $\mathbf{z} = \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\eta}$ has an error term with a covariance matrix $\sigma^2\mathbf{I}_N$. This is the form of a standard linear model, so we can use ordinary least squares (OLS) to estimate $\boldsymbol{\beta}$.

We minimize the sum of squared errors, $S(\boldsymbol{\beta})$:
$$S(\boldsymbol{\beta}) = (\mathbf{z} - \mathbf{B}\boldsymbol{\beta})'(\mathbf{z} - \mathbf{B}\boldsymbol{\beta})$$

The OLS estimator that minimizes this sum is given by the normal equations:
$$\hat{\boldsymbol{\beta}} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{z}$$

Now, we substitute back the definitions of $\mathbf{B}$ and $\mathbf{z}$: * $\mathbf{B}'\mathbf{B} = (\mathbf{V}^{-1/2})'(\mathbf{V}^{-1/2}) = \mathbf{V}^{-1/2}\mathbf{V}^{-1/2} = \mathbf{V}^{-1}$

- $(\mathbf{B}'\mathbf{B})^{-1} = (\mathbf{V}^{-1})^{-1} = \mathbf{V}$
- $\mathbf{z} = \mathbf{B}\mathbf{y} = \mathbf{V}^{-1/2}\mathbf{y}$

Substituting these into the expression for $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = \mathbf{V}(\mathbf{V}^{-1/2})'(\mathbf{V}^{-1/2}\mathbf{y}) = \mathbf{V}(\mathbf{V}^{-1/2}\mathbf{V}^{-1/2})\mathbf{y} = \mathbf{V}\mathbf{V}^{-1}\mathbf{y} = \mathbf{I}_N\mathbf{y} = \mathbf{y}$$

The unconstrained least squares estimate of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}} = \mathbf{y}$$

2.0.3 3. Constrained Least Squares Estimate

Now we obtain the estimate of $\boldsymbol{\beta}$ under the linear constraint $\mathbf{C}'\boldsymbol{\beta} = \mathbf{0}$. We use the method of Lagrange multipliers to minimize the sum of squares subject to this constraint.

The objective function to minimize is the generalized sum of squares, which is equivalent to the sum of squares from the transformed model:

$$S(\beta) = (\mathbf{y} - \beta)'\mathbf{V}^{-1}(\mathbf{y} - \beta)$$

The Lagrangian function $\mathcal{L}$ is:

$$\mathcal{L}(\beta, \boldsymbol{\lambda}) = (\mathbf{y} - \beta)'\mathbf{V}^{-1}(\mathbf{y} - \beta) + 2\boldsymbol{\lambda}'(\mathbf{C}'\beta)$$

where $2\boldsymbol{\lambda}$ is the vector of Lagrange multipliers.

Taking the partial derivative with respect to $\boldsymbol{\beta}$ and setting it to zero gives:

$$\frac{\partial \mathcal{L}}{\partial \beta} = -2\mathbf{V}^{-1}(\mathbf{y} - \beta) + 2\mathbf{C}\boldsymbol{\lambda} = \mathbf{0}$$

$$\mathbf{V}^{-1}(\mathbf{y} - \hat{\boldsymbol{\beta}}_c) = \mathbf{C}\boldsymbol{\lambda}$$

Solving for the constrained estimate $\hat{\boldsymbol{\beta}}_c$:

$$\mathbf{y} - \hat{\boldsymbol{\beta}}_c = \mathbf{V}\mathbf{C}\boldsymbol{\lambda}$$
$$\hat{\boldsymbol{\beta}}_c = \mathbf{y} - \mathbf{V}\mathbf{C}\boldsymbol{\lambda}$$

To find $\boldsymbol{\lambda}$, we apply the constraint $\mathbf{C}'\hat{\boldsymbol{\beta}}_c = \mathbf{0}$:

$$\mathbf{C}'(\mathbf{y} - \mathbf{V}\mathbf{C}\boldsymbol{\lambda}) = \mathbf{0}$$

$$\mathbf{C}'\mathbf{y} - \mathbf{C}'\mathbf{V}\mathbf{C}\boldsymbol{\lambda} = \mathbf{0}$$
$$\mathbf{C}'\mathbf{V}\mathbf{C}\boldsymbol{\lambda} = \mathbf{C}'\mathbf{y}$$

Solving for $\boldsymbol{\lambda}$:

$$\boldsymbol{\lambda} = (\mathbf{C}'\mathbf{V}\mathbf{C})^{-1}\mathbf{C}'\mathbf{y}$$

Finally, substitute this expression for $\boldsymbol{\lambda}$ back into the equation for $\hat{\boldsymbol{\beta}}_c$:

$$\hat{\boldsymbol{\beta}}_c = \mathbf{y} - \mathbf{V}\mathbf{C}(\mathbf{C}'\mathbf{V}\mathbf{C})^{-1}\mathbf{C}'\mathbf{y}$$

This is the constrained least squares estimate of $\boldsymbol{\beta}$.

3 3

4.27. Consider the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix},$$

and $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$. If possible, obtain the b.l.u.e.'s of (i) $\beta_1 + \beta_3$, and (ii) β_2, and compute their variances.

3.1 Step-by-Step Derivations

3.1.1 1. The Corrected Model Matrix

Since the 4×3 design matrix $\mathbf{X}$ is:

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

3.1.2 2. Checking the Model Rank

To find the b.l.u.e., we first check the rank of the model. The columns of $\mathbf{X}$ are:

$$c_1 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \quad c_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \quad c_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 1 \end{pmatrix}$$

We can see that **column 1 is identical to column 3** ($c_1 = c_3$). This means the columns are linearly dependent, and the matrix $\mathbf{X}$ is not of full rank (its rank is 2, not 3). Consequently, the matrix $\mathbf{X}^T\mathbf{X}$ is not invertible, and we cannot estimate β_1, β_2, and β_3 individually.

3.1.3 3. Reparameterization of the Model

Since $c_1 = c_3$, the model equation can be rewritten:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon = c_1\beta_1 + c_2\beta_2 + c_3\beta_3 + \epsilon$$

$$\mathbf{y} = c_1\beta_1 + c_2\beta_2 + c_1\beta_3 + \epsilon$$
$$\mathbf{y} = c_1(\beta_1 + \beta_3) + c_2\beta_2 + \epsilon$$

This shows we can estimate the combinations $\gamma_1 = \beta_1 + \beta_3$ and $\gamma_2 = \beta_2$. We can define a new full-rank model $\mathbf{y} = \mathbf{Z}\gamma + \epsilon$, where:

$$\mathbf{Z} = \begin{pmatrix} c_1 & c_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_3 \\ \beta_2 \end{pmatrix}$$

3.1.4 4. Finding the Estimators

Now we can use the standard ordinary least squares (OLS) formula for the full-rank model with matrix $\mathbf{Z}$. The b.l.u.e. of γ is $\hat{\gamma} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$.

First, calculate $\mathbf{Z}^T\mathbf{Z}$:

$$\mathbf{Z}^T\mathbf{Z} = \begin{pmatrix} 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ -1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}$$

Next, find its inverse:

$$(\mathbf{Z}^T\mathbf{Z})^{-1} = \frac{1}{3 \cdot 3 - (-2)(-2)} \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$$

Now, calculate $\mathbf{Z}^T\mathbf{y}$:

$$\mathbf{Z}^T\mathbf{y} = \begin{pmatrix} 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} = \begin{pmatrix} Y_2 - Y_3 + Y_4 \\ Y_1 + Y_3 - Y_4 \end{pmatrix}$$

Finally, compute $\hat{\gamma}$:

$$\hat{\gamma} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} Y_2 - Y_3 + Y_4 \\ Y_1 + Y_3 - Y_4 \end{pmatrix}$$

$$\hat{\gamma} = \frac{1}{5} \begin{pmatrix} 3(Y_2 - Y_3 + Y_4) + 2(Y_1 + Y_3 - Y_4) \\ 2(Y_2 - Y_3 + Y_4) + 3(Y_1 + Y_3 - Y_4) \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 2Y_1 + 3Y_2 - Y_3 + Y_4 \\ 3Y_1 + 2Y_2 + Y_3 - Y_4 \end{pmatrix}$$

3.1.5 5. Part (i): Solution for $\beta_1 + \beta_3$

The b.l.u.e. of $\beta_1 + \beta_3$ is $\hat{\gamma}_1$.

$$\widehat{\beta_1 + \beta_3} = \frac{1}{5}(2Y_1 + 3Y_2 - Y_3 + Y_4)$$

The variance of the estimators is given by $Var(\hat{\gamma}) = (\mathbf{Z}^T\mathbf{Z})^{-1}\sigma^2$. The variance of $\hat{\gamma}_1$ is the first diagonal element of this matrix.

$$Var(\widehat{\beta_1 + \beta_3}) = Var(\hat{\gamma}_1) = \frac{3}{5}\sigma^2$$

3.1.6 6. Part (ii): Solution for β_2

The b.l.u.e. of β_2 is $\hat{\gamma}_2$.

$$\hat{\beta}_2 = \frac{1}{5}(3Y_1 + 2Y_2 + Y_3 - Y_4)$$

The variance of $\hat{\beta}_2$ is the second diagonal element of the variance-covariance matrix.

$$Var(\hat{\beta}_2) = Var(\hat{\gamma}_2) = \frac{3}{5}\sigma^2$$

4 4

We refer to $\boldsymbol{\beta}_r^0$ in (4.6.4) as the constrained LS solution of $\boldsymbol{\beta}$ in the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ subject to $\mathbf{A}'\boldsymbol{\beta} = \mathbf{b}$. Let

$$\widehat{\mathbf{y}}_r = \mathbf{X}\boldsymbol{\beta}_r^0, \quad \widehat{\boldsymbol{\varepsilon}}_r = \mathbf{y} - \widehat{\mathbf{y}}_r, \quad \text{and} \quad SSE_r = \widehat{\boldsymbol{\varepsilon}}_r'\widehat{\boldsymbol{\varepsilon}}_r \tag{4.6.5}$$

be the fitted vector, residual vector and error sum of squares in the restricted model.

4.28. For the constrained estimates (4.6.5), show that $SSE_r = SSE + \|\widehat{\mathbf{y}} - \widehat{\mathbf{y}}_r\|^2$.

This identity is proven by decomposing the restricted sum of squares (SSE_r) into the unrestricted sum of squares (SSE) and the sum of squares due to the difference between the fitted values. The proof relies on the fundamental property that the ordinary least squares (OLS) residual vector is orthogonal to the column space of the design matrix $\mathbf{X}$.

4.0.1 Derivation of the Identity

We want to show that $SSE_r = SSE + ||\hat{\mathbf{y}} - \hat{\mathbf{y}}_r||^2$.

1. Start with the Definition of SSE_r

The restricted sum of squared errors is defined as the squared norm of the restricted residual vector:

$$SSE_r = ||\mathbf{y} - \hat{\mathbf{y}}_r||^2$$

2. Introduce the Unrestricted Fitted Vector $\hat{\mathbf{y}}$

We can add and subtract the unrestricted fitted vector, $\hat{\mathbf{y}}$, inside the norm without changing the value:

$$SSE_r = ||(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \hat{\mathbf{y}}_r)||^2$$

3. Expand the Squared Norm

Expanding the squared norm (which is the inner product of the vector with itself) gives three terms:

$$SSE_r = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \hat{\mathbf{y}}_r)'(\hat{\mathbf{y}} - \hat{\mathbf{y}}_r) + 2(\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \hat{\mathbf{y}}_r)$$

This can be rewritten using norm notation:

$$SSE_r = ||\mathbf{y} - \hat{\mathbf{y}}||^2 + ||\hat{\mathbf{y}} - \hat{\mathbf{y}}_r||^2 + 2(\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \hat{\mathbf{y}}_r)$$

The first term is the definition of the unrestricted sum of squares, SSE. So, our equation becomes:

$$SSE_r = SSE + ||\hat{\mathbf{y}} - \hat{\mathbf{y}}_r||^2 + 2(\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \hat{\mathbf{y}}_r)$$

4. Show the Cross-Product Term is Zero

The proof is complete if we can show that the cross-product term, $2(\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \hat{\mathbf{y}}_r)$, is equal to zero.

- The vector $(\mathbf{y} - \hat{\mathbf{y}})$ is the OLS residual vector, $\hat{\boldsymbol{\varepsilon}}$.
- The vectors $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{y}}_r = \mathbf{X}\hat{\boldsymbol{\beta}}_r$ are both linear combinations of the columns of $\mathbf{X}$. Therefore, they both lie in the column space of $\mathbf{X}$, denoted $C(\mathbf{X})$. Their difference, $(\hat{\mathbf{y}} - \hat{\mathbf{y}}_r)$, must also lie in $C(\mathbf{X})$.

A fundamental property of ordinary least squares is that the residual vector $\hat{\boldsymbol{\varepsilon}} = (\mathbf{y} - \hat{\mathbf{y}})$ is **orthogonal** to every vector in the column space $C(\mathbf{X})$. This means their inner product is zero.

Therefore, the inner product of the residual vector $(\mathbf{y} - \hat{\mathbf{y}})$ and the vector $(\hat{\mathbf{y}} - \hat{\mathbf{y}}_r)$ is zero:

$$(\mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} - \hat{\mathbf{y}}_r) = 0$$

5. Final Result

Substituting this result back into our expanded equation, we get:

$$SSE_r = SSE + ||\hat{\mathbf{y}} - \hat{\mathbf{y}}_r||^2 + 2(0)$$

$$\implies SSE_r = SSE + ||\hat{\mathbf{y}} - \hat{\mathbf{y}}_r||^2$$

This completes the proof. This relationship can be interpreted as a geometric result similar to the **Pythagorean theorem** in vector space, where the total squared length (SSE_r) is the sum of the squared lengths of two orthogonal components.

5 5

Example 4.6.2. Let $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, with $\mathrm{E}(\varepsilon_{ij}) = 0$, $i = 1, \cdots, a$, $j = 1, \cdots, n$, and suppose that we impose the model constraint $\sum_{i=1}^{a} c_i\tau_i = 0$, with $\sum_{i=1}^{a} c_i \neq 0$. We obtain the LS solutions for μ and $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_a)'$ using the method of Lagrangian multipliers. Let

$$S_r(\mu, \boldsymbol{\tau}, \lambda) = \sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij} - \mu - \tau_i)^2 + 2\lambda\sum_{i=1}^{a} c_i\tau_i.$$

Then we need to solve

$$\frac{\partial S}{\partial \mu} = -2\sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij} - \mu - \tau_i) = 0,$$

$$\frac{\partial S}{\partial \tau_i} = -2\sum_{j=1}^{n}(Y_{ij} - \mu - \tau_i) + 2\lambda c_i = 0, \quad 1 \le i \le a.$$

Adding the last a equations and subtracting the first one gives $2\lambda\sum_{i=1}^{a} c_i = 0$. Since $\sum_{i=1}^{a} c_i \neq 0$, then $\lambda = 0$. Thus $(\mu^0, \tau_1^0, \cdots, \tau_a^0)$ is a solution under the constraint if and only if $\tau_i^0 = \mu^0 - \overline{Y}_{i\cdot}$ and $\sum_{i=1}^{a} c_i\tau_i = 0$. It then follows that $\mu^0 = \sum_{i=1}^{a} c_i\overline{Y}_{i\cdot}/\sum_{i=1}^{a} c_i$ and $\tau_i^0 = \overline{Y}_{i\cdot} - \mu^0$, $i = 1, \cdots, a$, form the unique solution. The fitted values are $\widehat{Y}_{ij} = \mu^0 + \tau_i^0 = \overline{Y}_{i\cdot}$, which are the same as in the unconstrained model. Indeed, since $\sum_{i=1}^{a} c_i\tau_i = \mathbf{A}'\boldsymbol{\beta}$, where $\mathbf{A} = (0, c_1, \cdots, c_a)'$ and $\boldsymbol{\beta} = (\mu, \tau_1, \cdots, \tau_a)'$, it is easy to see that $\mathcal{C}(\mathbf{A}) \cap \mathcal{R}(\mathbf{X}) = \{\mathbf{0}\}$. Then by Result 4.6.5, the fitted values are the same under the model constraint as those in the unconstrained case. □

4.30. Continuing Example 4.6.2, suppose we impose the constraints $\mu = s$ and $\tau_1 + \tau_2 = t$, where s and t are constants.

(a) Show that μ and $\tau_1 + \tau_2$ are not estimable while $2\mu + \tau_1 + \tau_2$ is.

(b) Show that the vector of fitted values under the above constraints is identical to the one under the constraint $2\mu + \tau_1 + \tau_2 = 2s + t$.

The model from Example 4.6.2 is the one-way analysis of variance (ANOVA) model:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad \text{for } i = 1, \ldots, a;\ j = 1, \ldots, n$$

The parameter vector is $\boldsymbol{\beta}' = (\mu, \tau_1, \ldots, \tau_a)$. This model is not of full rank because the first column of the design matrix $\mathbf{X}$ (a vector of all ones) is the sum of the other a columns (which are indicators for each group).

5.0.1 (a) Estimability

A linear combination of parameters $\mathbf{c}'\boldsymbol{\beta}$ is **estimable** if and only if the vector $\mathbf{c}$ can be expressed as a linear combination of the rows of the design matrix $\mathbf{X}$. For the one-way ANOVA model, this is equivalent to the condition that the coefficient of μ must equal the sum of the coefficients of the τ_i's. Let $\mathbf{c}'\boldsymbol{\beta} = c_0\mu + \sum_{i=1}^{a} c_i\tau_i$. The combination is estimable if and only if $c_0 = \sum_{i=1}^{a} c_i$.

Check the given linear combinations. The problem involves τ_1 and τ_2, so we assume there are at least two groups ($a \geq 2$).

1. **For μ:** The combination is $1 \cdot \mu + 0 \cdot \tau_1 + 0 \cdot \tau_2 + \ldots$.
   - Here, $c_0 = 1$ and $\sum c_i = 0$.
   - Since $1 \neq 0$, the condition is not met. Thus, μ **is not estimable.**
2. **For $\tau_1 + \tau_2$:** The combination is $0 \cdot \mu + 1 \cdot \tau_1 + 1 \cdot \tau_2 + 0 \cdot \tau_3 + \ldots$.
   - Here, $c_0 = 0$ and $\sum c_i = 1 + 1 = 2$.
   - Since $0 \neq 2$, the condition is not met. Thus, $\tau_1 + \tau_2$ **is not estimable.**
3. **For $2\mu + \tau_1 + \tau_2$:** The combination is $2 \cdot \mu + 1 \cdot \tau_1 + 1 \cdot \tau_2 + 0 \cdot \tau_3 + \ldots$.
   - Here, $c_0 = 2$ and $\sum c_i = 1 + 1 = 2$.
   - Since $2 = 2$, the condition is met. Thus, $2\mu + \tau_1 + \tau_2$ **is estimable.**

5.0.2 (b) Equivalence of Fitted Values

The vector of fitted values under constraints, $\hat{\mathbf{y}}_r$, is the orthogonal projection of the data vector $\mathbf{y}$ onto the subspace of possible mean vectors ($\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$) that satisfy those constraints. To show that the fitted values are identical under two different sets of constraints, we must show that both sets define the **exact same subspace** of mean vectors.

Let S_1 be the subspace defined by the constraints $\{\mu = s,\ \tau_1 + \tau_2 = t\}$, and let S_2 be the subspace defined by the constraint $\{2\mu + \tau_1 + \tau_2 = 2s + t\}$. We need to prove that $S_1 = S_2$.

Proof that $S_1 \subseteq S_2$ Let $\boldsymbol{\eta}$ be any vector in S_1. This means $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ for some parameter vector $\boldsymbol{\beta}$ that satisfies $\mu = s$ and $\tau_1 + \tau_2 = t$. If we check the constraint for S_2 using these parameters, we get:

$$2\mu + \tau_1 + \tau_2 = 2(s) + (t) = 2s + t$$

The constraint for S_2 is satisfied. Therefore, any mean vector in S_1 is also in S_2.

Proof that $S_2 \subseteq S_1$ Let $\boldsymbol{\eta}$ be any vector in S_2. This means $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ for some $\boldsymbol{\beta}$ where $2\mu + \tau_1 + \tau_2 = 2s + t$. Because the model is not of full rank, there are infinitely many parameter vectors that produce the same mean vector $\boldsymbol{\eta}$. We can find a new parameter vector $\boldsymbol{\beta}^* = \boldsymbol{\beta} + k\mathbf{d}$, where $\mathbf{d}$ is a vector in the **null space** of $\mathbf{X}$ (meaning $\mathbf{Xd} = \mathbf{0}$), such that $\mathbf{X}\boldsymbol{\beta}^* = \boldsymbol{\eta}$. For the one-way ANOVA model, vectors in the null space have the form $\mathbf{d}' = (1, -1, -1, \ldots, -1)$. Let's define a new parameter vector $\boldsymbol{\beta}^* = (\mu^*, \tau_1^*, \ldots, \tau_a^*)$ where:

- $\mu^* = \mu + k$
- $\tau_i^* = \tau_i - k$ for all i

We need to find a scalar k such that $\boldsymbol{\beta}^*$ satisfies the constraints for S_1, namely $\mu^* = s$ and $\tau_1^* + \tau_2^* = t$. From the first constraint:

$$\mu^* = s \implies \mu + k = s \implies k = s - \mu$$

Now, we check if this choice of k makes the second constraint true:

$$\tau_1^* + \tau_2^* = (\tau_1 - k) + (\tau_2 - k) = \tau_1 + \tau_2 - 2k$$

Substitute $k = s - \mu$:

$$\tau_1 + \tau_2 - 2(s - \mu) = \tau_1 + \tau_2 - 2s + 2\mu$$

Since the original $\boldsymbol{\beta}$ satisfied the constraint for S_2, we know that $2\mu + \tau_1 + \tau_2 = 2s + t$. Rearranging this gives $2\mu + \tau_1 + \tau_2 - 2s = t$. Substituting this into our expression:

$$\tau_1^* + \tau_2^* = (2\mu + \tau_1 + \tau_2 - 2s) = t$$

The second constraint is also satisfied. This shows that for any mean vector in S_2, we can find a parameterization that satisfies the constraints for S_1. Therefore, any mean vector in S_2 is also in S_1.

Conclusion Since $S_1 \subseteq S_2$ and $S_2 \subseteq S_1$, the two sets are identical ($S_1 = S_2$). Because both sets of constraints define the same subspace of mean vectors, the orthogonal projection of $\mathbf{y}$ onto this subspace—the vector of fitted values—must be identical.

6 6

5.4. Let X_1 and X_2 be random variables such that $X_1 + X_2$ and $X_1 - X_2$ have independent standard normal distributions. Show that $(X_1, X_2)'$ has a bivariate normal distribution.

This can be shown by defining a new vector of random variables and using the property that a linear transformation of a multivariate normal random vector is also multivariate normal.

6.0.1 Proof

1. Define New Variables and Their Distribution

Let's define two new random variables: * $Y_1 = X_1 + X_2$

- $Y_2 = X_1 - X_2$

We are given that Y_1 and Y_2 have independent standard normal distributions. This means:

- $Y_1 \sim N(0, 1)$
- $Y_2 \sim N(0, 1)$
- Y_1 and Y_2 are independent.

Because Y_1 and Y_2 are independent normal random variables, their joint distribution is a bivariate normal distribution. Let $\mathbf{Y} = (Y_1, Y_2)'$. The distribution of $\mathbf{Y}$ is:

$$\mathbf{Y} \sim N(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$$

where the mean vector is $\boldsymbol{\mu}_Y = \begin{pmatrix} E[Y_1] \\ E[Y_2] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and the covariance matrix is $\boldsymbol{\Sigma}_Y = \begin{pmatrix} \text{Var}(Y_1) & \text{Cov}(Y_1, Y_2) \\ \text{Cov}(Y_1, Y_2) & \text{Var}(Y_2) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}_2$, since they are independent.

2. Express X_1 and X_2 as a Linear Transformation

Next, we express the original variables X_1 and X_2 in terms of Y_1 and Y_2. By solving the system of linear equations: * Adding the two equations: $Y_1 + Y_2 = (X_1 + X_2) + (X_1 - X_2) = 2X_1 \implies X_1 = \frac{1}{2}Y_1 + \frac{1}{2}Y_2$ * Subtracting the second equation from the first: $Y_1 - Y_2 = (X_1 + X_2) - (X_1 - X_2) = 2X_2 \implies X_2 = \frac{1}{2}Y_1 - \frac{1}{2}Y_2$

This linear transformation can be written in matrix form. Let $\mathbf{X} = (X_1, X_2)'$:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

This is a linear transformation of the form $\mathbf{X} = \mathbf{AY}$, where $\mathbf{A} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}$.

3. Apply the Property of Linear Transformations

A key property of the multivariate normal distribution is that any linear transformation of a multivariate normal random vector results in another multivariate normal random vector.

Since $\mathbf{Y}$ has a bivariate normal distribution and $\mathbf{X}$ is a linear transformation of $\mathbf{Y}$, it follows directly that $\mathbf{X} = (X_1, X_2)'$ **must also have a bivariate normal distribution.**

6.0.2 Distribution Parameters (Optional)

We can also find the parameters of the distribution of $\mathbf{X}$.

- **Mean Vector:**
$$E[\mathbf{X}] = E[\mathbf{AY}] = \mathbf{A}E[\mathbf{Y}] = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- **Covariance Matrix:**

$$\text{Cov}(\mathbf{X}) = \text{Cov}(\mathbf{AY}) = \mathbf{A}\text{Cov}(\mathbf{Y})\mathbf{A}' = \mathbf{A}\mathbf{I}_2\mathbf{A}' = \mathbf{AA}'$$

$$\text{Cov}(\mathbf{X}) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} = \begin{pmatrix} 1/4 + 1/4 & 1/4 - 1/4 \\ 1/4 - 1/4 & 1/4 + 1/4 \end{pmatrix} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

Thus, $(X_1, X_2)'$ has a bivariate normal distribution with a mean of $\mathbf{0}$ and a covariance matrix of $\frac{1}{2}\mathbf{I}_2$. This also implies that X_1 and X_2 are independent normal random variables, each with a distribution of $N(0, 1/2)$.