Lab 12

2024-12-08

Q1

For the models below, determine whether the OLS estimator of the parameter vector is (i) unique, (ii) identifiable, (iii) estimable. (iv) Briefly justify your response.

(a)

One-Way Cell Means ANOVA

Uniqueness: Not Unique. The design matrix is not full rank due to the unconstrained group means.

Identifiability: Not Identifiable. Parameters μ_i are not identifiable without constraints (e.g., sum-to-zero).

Estimability: Partially. Contrasts (e.g., $\mu_i - \mu_j$) are estimable, but individual μ_i are not.

(b)

One-Way Effects ANOVA

Uniqueness: Not Unique. The design matrix is not full rank due to the inclusion of an intercept alongside group indicator variables.

Identifiability: Not Identifiable. Group effects α_i are not identifiable without constraints (e.g., sum-to-zero or setting one group effect to zero).

Estimability: Partially. Contrasts (e.g., $\alpha_i - \alpha_j$) are estimable, but individual group effects are not.

(c)

Randomized Complete Block Design (RCBD)

Uniqueness: Not Unique. The design matrix is not full rank due to the inclusion of both block and treatment indicator variables without constraints.

Identifiability: Not Identifiable. Treatment effects τ_i and block effects β_j are not identifiable without constraints (e.g., sum-to-zero for treatments and blocks).

Estimability: Partially. Contrasts of treatment effects (e.g., $\tau_i - \tau_k$) and block effects (e.g., $\beta_j - \beta_l$) are estimable, but individual effects are not.

(d)

Two-Way Effects ANOVA

Uniqueness: Not Unique. The design matrix is not full rank due to the inclusion of an intercept, main effects, and interaction terms without constraints.

Identifiability: Not Identifiable. Main effects α_i, β_j, and interaction effects $(\alpha\beta)_{ij}$ are not identifiable without constraints (e.g., sum-to-zero for main effects and interactions).

Estimability: Partially. Contrasts of main effects (e.g., $\alpha_i - \alpha_k$, $\beta_j - \beta_l$) and interaction effects (e.g., $(\alpha\beta)_{ij} - (\alpha\beta)_{kl}$) are estimable, but individual effects are not.

(e)

Two-Way Cell Means ANOVA

Uniqueness: Not Unique. The design matrix is not full rank due to the unconstrained group means for each combination of factors.

Identifiability: Not Identifiable. Cell means μ_{ij} are not identifiable without constraints (e.g., setting one cell mean to zero or sum-to-zero constraints).

Estimability: Partially. Contrasts of cell means (e.g., $\mu_{ij} - \mu_{kl}$) are estimable, but individual cell means are not.

Q2

A completely randomized two-factor experiment described by Hunter (1989) consisted of burning fuel with levels of two additives in a laboratory setting and determining the CO (carbon monoxide) emissions released. Eighteen batches of a standard fuel were available for this study. Two of the batches were randomly assigned to each of nine combinations of two additives corresponding to three levels of added ethanol (0.1, 0.2, or 0.3) and three air/fuel ratio settings (14, 15, or 16). Units for the ethanol levels were not reported. CO emissions concentrations (g/meter^3) were determined for each burning the same amount of fuel from each of the 18 batches. The data are shown below.

Added Ethanol	Air/Fuel Ratio 14	Air/Fuel Ratio 15	Air/Fuel Ratio 16
0.1	66, 62	72, 67	68, 66
0.2	78, 81	80, 81	66, 69
0.3	90, 94	75, 78	60, 58

Consider the model:

$$Y_{ijk} = \mu + \alpha_i + \tau_j + (\alpha\tau)_{ij} + \epsilon_{ijk}$$

Where α_i represents the i-th level of added ethanol effect, τ_j represents the j-th level of air/fuel ratio effect, and k denotes the replicates.

(a)

Show this model meets the definition of a linear model by writing the design matrix X and the parameter vector β. To save time and room, only write the unique rows of the design matrix X.

$$X = (\mu, \alpha_1, \alpha_2, \alpha_3, \tau_1, \tau_2, \tau_3, (\alpha\tau)_{11}, (\alpha\tau)_{12}, (\alpha\tau)_{13}, (\alpha\tau)_{21}, (\alpha\tau)_{22}, (\alpha\tau)_{23}, (\alpha\tau)_{31}, (\alpha\tau)_{32}, (\alpha\tau)_{33})$$

I needed to actually write out and then knit as an image the design matrix. The size of it was not allowing me to knit in environment. So apologies if the design matrix is located in a weird spot of my output.

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 1: CocoMelon

$$\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ (\alpha\tau)_{11} \\ (\alpha\tau)_{12} \\ (\alpha\tau)_{13} \\ (\alpha\tau)_{21} \\ (\alpha\tau)_{22} \\ (\alpha\tau)_{23} \\ (\alpha\tau)_{31} \\ (\alpha\tau)_{32} \\ (\alpha\tau)_{33} \end{bmatrix}$$

$$\mathbf{X}\beta = \begin{bmatrix} \mu + \alpha_1 + \tau_1 + (\alpha\tau)_{11} \\ \mu + \alpha_1 + \tau_2 + (\alpha\tau)_{12} \\ \mu + \alpha_1 + \tau_3 + (\alpha\tau)_{13} \\ \mu + \alpha_2 + \tau_1 + (\alpha\tau)_{21} \\ \mu + \alpha_2 + \tau_2 + (\alpha\tau)_{22} \\ \mu + \alpha_2 + \tau_3 + (\alpha\tau)_{23} \\ \mu + \alpha_3 + \tau_1 + (\alpha\tau)_{31} \\ \mu + \alpha_3 + \tau_2 + (\alpha\tau)_{32} \\ \mu + \alpha_3 + \tau_3 + (\alpha\tau)_{33} \end{bmatrix}$$

So we do satisfy the equation for the general linear model:

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\epsilon}$$

(b)

What extra assumption must be added to the linear model in part (a) for the model to be considered a Gauss-Markov linear model?

We have additional assumptions regarding the error term, $\boldsymbol{\epsilon}$, specifically 1. Constant variance: $\mathrm{Var}(\epsilon_{ijk}) = \sigma^2$ for all (i, j, k) 2. Errors uncorrelated: $\mathrm{Cov}(\epsilon_{ijk}, \epsilon_{lmn}) = 0$ for all $(i, j, k) \neq (l, m, n)$

(c)

For the Gauss-Markov model, list the conditions under which $Y^T(I - P_X)Y$ has a quadratic form.

We need the following: - 1. Linearity of the Model - 2. Full Rank of the Design Matrix - 3. Orthogonal Projection, i.e. orthogonal $(I - P_X)$ - 4.Constant Variance of Errors, from part (b) - 5. Uncorrelated Errors, from part (b)

(d)

For the Normal-theory Gauss-Markov model, explain why the value $\frac{Y^T(I-P_X)Y}{\sigma^2}$ has a Central Chi-square distribution with 9 degrees of freedom.

$Y^T(I - P_X)Y$ is a quadratic form of the normal vector ϵ.

$I - P_X$ is symmetric, idempotent, and has rank 9, providing the 9 degrees of freedom for its associated distribution.

The residuals are orthogonal to the column space of X.

Dividing by σ^2 standardizes the variance, yielding the χ^2 distribution with 9 degrees of freedom.

(e)

Show the function $\alpha_1 + \alpha_2$ is not estimable.

Under the "sum-to-zero" constraint, we know:

$$\alpha_1 + \alpha_2 + \alpha_3 = 0 \rightarrow \alpha_1 + \alpha_2 = -\alpha_3$$

Since $\alpha_1 + \alpha_2$ depends on α_3, it cannot be expressed solely in terms of the observed data. This makes $\alpha_1 + \alpha_2$ unidentifiable from the data and hence not estimable.

The linear combination $\alpha_1 + \alpha_2$ cannot be written as a linear combination of the rows of X, as the columns corresponding to α_1, α_2, and α_3 are linearly dependent due to the sum-to-zero constraint.

(f)

Show the function $\tau_2 - \tau_3$ is not estimable.

Similar to part (e), the "sum-to-zero" constraint on the tau's means:

$$\sum_{j=1}^{3} \tau_j = 0 \rightarrow \tau_1 + \tau_2 + \tau_3 = 0 \rightarrow \tau_2 - \tau_3 = -\tau_1 - 2\tau_3$$

Since $\tau_2 - \tau_3$ depends on τ_1, it cannot be determined uniquely from the observed data. This dependence prevents $\tau_2 - \tau_3$ from being identifiable and hence estimable.

The columns of X corresponding to τ_1, τ_2, and τ_3 are linearly dependent because of the sum-to-zero constraint. This prevents the linear combination $\tau_2 - \tau_3$ from being expressed as a linear combination of the rows of X.

(g)

Show the function $(\tau_1 - \tau_2) + [(\alpha\tau)_{11} - (\alpha\tau)_{12}]$ is estimable.

The main effects $\tau_1 - \tau_2$ represent a contrast between air/fuel ratio levels, which is estimable because contrasts of main effects are not affected by the sum-to-zero constraint on τ_j.

The interaction effect $(\alpha\tau)_{11} - (\alpha\tau)_{12}$ represents a contrast within the interaction terms for a fixed ethanol level (α_1) and two different air/fuel ratio levels (τ_1 and τ_2). Contrasts of interaction terms are also estimable under the sum-to-zero constraints on interactions.

Both $\tau_1 - \tau_2$ and $(\alpha\tau)_{11} - (\alpha\tau)_{12}$ can be expressed as linear combinations of the rows of the design matrix X.

The sum of two estimable functions, $(\tau_1 - \tau_2) + [(\alpha\tau)_{11} - (\alpha\tau)_{12}]$, is therefore also estimable.

The combination $(\tau_1 - \tau_2) + [(\alpha\tau)_{11} - (\alpha\tau)_{12}]$ lies entirely in the row space of the design matrix X, satisfying the condition for estimability.