STAT 5000

Statistical Methods I

Week 15

Fall 2024

Dr. Danica Ommen

Unit 3

Linear Model Theory:

Introduction

Linear Models

Linear models provide a unified approach to many models

- One-way ANOVA (including two-independent samples)
- Block designs with fixed blocks (including matched pairs)
- Two-way ANOVA
- Simple Linear Regression
- Multiple Linear Regression

Any linear model can be written in the form

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

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{\substack{\uparrow \\ \text{response vector}}} = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}}_{\substack{\uparrow \\ \text{the elements of design matrix } X \text{ are known (non-random) values}}} \underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}_{\substack{\uparrow \\ \text{unknown parameters}}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\substack{\uparrow \\ \text{random errors (not observed)}}}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \text{ is a random vector}$$

1. $E(\mathbf{Y}) = X\beta$ is a vector of expected responses for some known matrix X of constants and unknown parameter vector β
2. $Var(\mathbf{Y}) = \Sigma$
3. Complete the model by specifying a probability distribution for the possible values of $\mathbf{Y}$ or ϵ

Gauss-Markov Model

The linear model $\mathbf{Y} = X\beta + \epsilon$ is called a Gauss-Markov model if

$$Var(\mathbf{Y}) = Var(\epsilon) = \sigma^2 I$$

for some unknown constant σ^2.

- The observations (and the random errors) are mutually uncorrelated
- Every observation (and every random error) has the same variance

Normal Theory Gauss-Markov Model

A normal theory Gauss-Markov model is a Gauss-Markov model where $\mathbf{Y}$ (and ϵ) has a multivariate normal distribution.

$$\mathbf{Y} \sim N(X\beta, \sigma^2 I) \text{ implying } \epsilon \sim N(\mathbf{0}, \sigma^2 I)$$

The additional assumption of a normal distribution is

1. not needed for most estimation results
2. used to create confidence intervals and perform tests of hypotheses
3. used to obtain distributions for test statistics

Regression

Example 1: Yield of a chemical process

- Response Variable = Yield (Y)
- Explanatory Variable 1 = Temperature (x_1)
- Explanatory Variable 2 = Time (x_2)
- $n = 5$ observations

Example 1: Yield of a chemical process

Yield (%) Y	Temperature (^{o}F) X_1	Time (hr) X_2
77	160	1
82	165	3
84	165	2
89	170	1
94	175	2

Regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \quad \text{for } i = 1, 2, 3, 4, 5$$

Example 1

Regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \quad \text{for } i = 1, 2, 3, 4, 5$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \\ 1 & x_{51} & x_{52} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

ANOVA Table

Variation	d.f.	Sums of Squares	Mean Square
Model	2	$\sum_{i=1}^{5}(\hat{Y}_i - \bar{Y})^2 = \mathbf{Y}^T\,(P_X - P_1)\,\mathbf{Y}$	$\frac{1}{2}\,SS_{\text{model}}$
Error	2	$\sum_{i=1}^{5}(Y_i - \hat{Y}_i)^2 = \mathbf{Y}^T\,(I - P_X)\,\mathbf{Y}$	$\frac{1}{2}\,SS_{\text{error}}$
Total	4	$\sum_{i=1}^{5}(Y_i - \bar{Y})^2 = \mathbf{Y}^T\,(I - P_1)\,\mathbf{Y}$	

where $\quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad$ and $\quad \hat{Y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i}$

$$P_X = X(X^TX)^{-1}X^T \quad \text{and} \quad P_1 = \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$$

Regression: ANOVA

The corrected total sum of squares is

$$SS_{\text{total}} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = (\mathbf{Y} - \bar{Y}\mathbf{1})^T (\mathbf{Y} - \bar{Y}\mathbf{1})$$

Note that

$$P_1\mathbf{Y} = \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T\mathbf{Y} = \mathbf{1}\frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}\mathbf{1}$$

and

$$(\mathbf{Y} - \bar{Y}\mathbf{1}) = \mathbf{Y} - P_1\mathbf{Y} = (I - P_1)\mathbf{Y}$$

Regression: ANOVA

Then

$$
\begin{aligned}
SS_{\text{total}} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= (\mathbf{Y} - \bar{Y}\mathbf{1})^T(\mathbf{Y} - \bar{Y}\mathbf{1}) \\
&= ((I - P_1)\mathbf{Y})^T(I - P_1)\mathbf{Y} \\
&= \mathbf{Y}^T(I - P_1)^T(I - P_1)\mathbf{Y} \\
&= \mathbf{Y}^T(I - P_1)(I - P_1)\mathbf{Y} \\
&= \mathbf{Y}^T(I - P_1)\mathbf{Y}
\end{aligned}
$$

because $(I - P_1)$ is a symmetric and idempotent matrix

Regression: ANOVA

$$SS_{\text{model}} = \sum_{i=1}^{5} (\hat{Y}_i - \bar{Y})^2 = (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1})^T (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1})$$

Note that

$$\hat{\mathbf{Y}} = X\mathbf{b} = X(X^T X)^{-1} X^T \mathbf{Y} = P_X \mathbf{Y}$$

and

$$\hat{\mathbf{Y}} - \bar{Y}\mathbf{1} = P_X \mathbf{Y} - P_1 \mathbf{Y} = (P_X - P_1)\mathbf{Y}$$

Regression: ANOVA

Then

$$\begin{aligned} SS_{\text{model}} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 &= (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1})^T(\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) \\ &= ((P_X - P_1)\mathbf{Y})^T(P_X - P_1)\mathbf{Y} \\ &= \mathbf{Y}^T(P_X - P_1)^T(P_X - P_1)\mathbf{Y} \\ &= \mathbf{Y}^T(P_X - P_1)(P_X - P_1)\mathbf{Y} \\ &= \mathbf{Y}^T(P_X - P_1)\mathbf{Y} \end{aligned}$$

because $(P_X - P_1)$ is a symmetric and idempotent matrix

Regression: ANOVA

Then

$$
\begin{aligned}
SS_{\text{error}} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 &= (\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}}) \\
&= ((I - P_X)\mathbf{Y})^T(I - P_X)\mathbf{Y} \\
&= \mathbf{Y}^T(I - P_X)^T(I - P_X)\mathbf{Y} \\
&= \mathbf{Y}^T(I - P_X)(I - P_X)\mathbf{Y} \\
&= \mathbf{Y}^T(I - P_X)\mathbf{Y}
\end{aligned}
$$

because $I - P_X$ is a symmetric and idempotent matrix

Regression: ANOVA

Partition the corrected total sum of squares:

$$\begin{aligned} SS_{\text{total}} &= \mathbf{Y}^T(I - P_1)\mathbf{Y} \\ &= \mathbf{Y}^T(I - P_X + P_X - P_1)\mathbf{Y} \\ &= \mathbf{Y}^T(I - P_X)\mathbf{Y} + \mathbf{Y}^T(P_X - P_1)\mathbf{Y} \\ &= SS_{\text{error}} + SS_{\text{model}} \end{aligned}$$

One-Way ANOVA

Example 2: Blood coagulation times (in seconds) for blood samples from 12 different rats. Each rat was fed one of three diets, with 4 rats per diet.

- Response Variable = Blood coagulation times (Y)
- Explanatory Variable = Diet (A, B, or C)
- $n = 12$ observations

One-Way ANOVA

Cell Means Model

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \end{bmatrix}$$

One-Way ANOVA

Effects Model

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \end{bmatrix}$$

Two-Way ANOVA

Example 3: A full factorial experiment

- Experimental Units - 8 plots of trees - 5 trees per plot.
- Response Variable = Percentage of apples with spots (Y)
- Explanatory Variable 1 = Variety of Apple (A or B)
- Explanatory Variable 2 = Fungicide use (new or old)
- $n = 8$ observations

Two-Way ANOVA

Cell Means Model

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}$$

Two-Way ANOVA

Effects Model

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \tau_1 \\ \tau_2 \\ (\alpha\tau)_{11} \\ (\alpha\tau)_{12} \\ (\alpha\tau)_{21} \\ (\alpha\tau)_{22} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}$$

Unit 3

Linear Model Theory:

Estimation

Ordinary Least Squares (OLS) Estimator

For a linear model with $E(\mathbf{Y}) = X\beta$, any vector $\mathbf{b}$ that minimizes the sum of squared residuals

$$\begin{aligned} Q(\mathbf{b}) &= \sum_{i=1}^{n} (Y_i - \mathbf{X}_i^T \mathbf{b})^2 \\ &= (\mathbf{Y} - X\mathbf{b})^T (\mathbf{Y} - X\mathbf{b}) \end{aligned}$$

is an ordinary least squares (OLS) estimator for β.

- In this definition X_i denotes a column vector constructed from the i^{th} row of the $n \times k$ model matrix X.
- The parameter vector β is a $k \times 1$ vector.

Normal Equations

For $j = 1, 2, \ldots, k$, solve the set of equations

$$0 = \frac{\partial Q(\mathbf{b})}{\partial b_j} = 2 \sum_{i=1}^{n} (Y_i - \mathbf{X_i}^T \mathbf{b}) X_{ij}$$

These equations are expressed in matrix form as

$$\begin{aligned} \mathbf{0} &= X^T(\mathbf{Y} - X\mathbf{b}) \\ &= X^T\mathbf{Y} - X^T X\mathbf{b} \end{aligned}$$

or

$$X^T X\mathbf{b} = X^T\mathbf{Y}$$

These are called the “normal” equations.

Uniqueness of OLS Estimator

If $X_{n\times k}$ has full column rank, rank(X) = k and

- X^TX is non-singular
- $(X^TX)^{-1}$ exists and is unique

This means we can solve the normal equations for **b** as:

$$\begin{aligned} X^TX\mathbf{b} &= X^T\mathbf{Y} \\ (X^TX)^{-1}(X^TX)\mathbf{b} &= (X^TX)^{-1}X^T\mathbf{Y} \\ \mathbf{b} &= (X^TX)^{-1}X^T\mathbf{Y} \end{aligned}$$

and **b** is unique.

Example: One-way ANOVA

Cell Means Model

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \end{bmatrix}$$

Example: One-way ANOVA

X is full rank: $\text{rank}(X) = 3$

$$X^TX = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \quad (X^TX)^{-1} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix}$$

$$X^T\mathbf{Y} = \begin{bmatrix} \sum_{j=1}^{4} Y_{1j} \\ \sum_{j=1}^{4} Y_{2j} \\ \sum_{j=1}^{4} Y_{3j} \end{bmatrix} \quad (X^TX)^{-1}X^T\mathbf{Y} = \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \end{bmatrix}$$

OLS Estimator

If $\text{rank}(X) < k$, then

- there are infinitely many solutions to the normal equations
- if $(X^T X)^-$ is a generalized inverse of $X^T X$, then

$$\mathbf{b} = (X^T X)^- X^T \mathbf{Y}$$

is one of the many solutions of the normal equations.

Generalized Inverse

For a given $m \times n$ matrix A, any $n \times m$ matrix G that satisfies

$$AGA = A$$

is a *generalized inverse* of A.

Comments

- We will use A^- to denote a generalized inverse of A.
- There may be infinitely many generalized inverses.
- If A is an $m \times m$ non-singular matrix, then $G = A^{-1}$ is the unique generalized inverse for A.

Example: One-way ANOVA

Effects model

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \end{bmatrix}$$

Example: One-way ANOVA

X is not full rank: $\text{rank}(X) = 3 < k = 4$

$$X^T X = \begin{bmatrix} n & n_1 & n_2 & n_3 \\ n_1 & n_1 & 0 & 0 \\ n_2 & 0 & n_2 & 0 \\ n_3 & 0 & 0 & n_3 \end{bmatrix} = \begin{bmatrix} 12 & 4 & 4 & 4 \\ 4 & 4 & 0 & 0 \\ 4 & 0 & 4 & 0 \\ 4 & 0 & 0 & 4 \end{bmatrix}$$

$$X^T \mathbf{Y} = \begin{bmatrix} n\bar{Y}_{..} \\ n_1\bar{Y}_{1.} \\ n_2\bar{Y}_{2.} \\ n_3\bar{Y}_{3.} \end{bmatrix} = \begin{bmatrix} 12\bar{Y}_{..} \\ 4\bar{Y}_{1.} \\ 4\bar{Y}_{2.} \\ 4\bar{Y}_{3.} \end{bmatrix}$$

Example: One-way ANOVA

Solution A: A generalized inverse for $X^T X$ is

$$(X^T X)^- = \begin{bmatrix} 0 & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} & \begin{bmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{bmatrix}^{-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{n_1} & 0 & 0 \\ 0 & 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & 0 & \frac{1}{n_3} \end{bmatrix}$$

and a solution to the normal equations is

$$\mathbf{b} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & n_1^{-1} & 0 & 0 \\ 0 & 0 & n_2^{-1} & 0 \\ 0 & 0 & 0 & n_3^{-1} \end{bmatrix} \begin{bmatrix} n\bar{Y}_{..} \\ n_1\bar{Y}_{1.} \\ n_2\bar{Y}_{2.} \\ n_3\bar{Y}_{3.} \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \end{bmatrix}$$

Example: One-way ANOVA

Solution B: Another generalized inverse for X^TX is

$$(X^TX)^- = \begin{bmatrix} \begin{bmatrix} n_. & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{bmatrix}^{-1} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & \quad 0 & \quad 0 \end{matrix} & 0 \end{bmatrix} = \frac{1}{n_3} \begin{bmatrix} 1 & -1 & -1 & 0 \\ -1 & \frac{n_1+n_3}{n_1} & 1 & 0 \\ -1 & 1 & \frac{n_2+n_3}{n_2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and a solution to the normal equations is

$$\mathbf{b} = \frac{1}{n_3} \begin{bmatrix} 1 & -1 & -1 & 0 \\ -1 & \frac{n_1+n_3}{n_1} & 1 & 0 \\ -1 & 1 & \frac{n_2+n_3}{n_2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} n\bar{Y}_{..} \\ n_1\bar{Y}_{1.} \\ n_2\bar{Y}_{2.} \\ n_3\bar{Y}_{3.} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{3.} \\ \bar{Y}_{1.} - \bar{Y}_{3.} \\ \bar{Y}_{2.} - \bar{Y}_{3.} \\ 0 \end{bmatrix}$$

Example: One-way ANOVA

Solution C: Another generalized inverse for X^TX is

$$(X^TX)^- = \frac{1}{n_1}\begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & \frac{n_1+n_2}{n_2} & 1 \\ -1 & 0 & 1 & \frac{n_1+n_3}{n_3} \end{bmatrix}$$

and a solution to the normal equations is

$$\mathbf{b} = \frac{1}{n_1}\begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & \frac{n_1+n_2}{n_2} & 1 \\ -1 & 0 & 1 & \frac{n_1+n_3}{n_3} \end{bmatrix}\begin{bmatrix} n\bar{Y}_{..} \\ n_1\bar{Y}_{1.} \\ n_2\bar{Y}_{2.} \\ n_3\bar{Y}_{3.} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{1.} \\ 0 \\ \bar{Y}_{2.} - \bar{Y}_{1.} \\ \bar{Y}_{3.} - \bar{Y}_{1.} \end{bmatrix}$$

Example: One-way ANOVA

Solution D: Another generalized inverse for X^TX is

$$(X^TX)^- = \begin{bmatrix} \frac{2}{n} & -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & \frac{1}{n_1} & 0 & 0 \\ -\frac{1}{n} & 0 & \frac{1}{n_2} & 0 \\ -\frac{1}{n} & 0 & 0 & \frac{1}{n_3} \end{bmatrix}$$

and a solution to the normal equations is

$$\mathbf{b} = \begin{bmatrix} \frac{2}{n} & -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & \frac{1}{n_1} & 0 & 0 \\ -\frac{1}{n} & 0 & \frac{1}{n_2} & 0 \\ -\frac{1}{n} & 0 & 0 & \frac{1}{n_3} \end{bmatrix} \begin{bmatrix} n\bar{Y}_{..} \\ n_1\bar{Y}_{1.} \\ n_2\bar{Y}_{2.} \\ n_3\bar{Y}_{3.} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{..} \\ \bar{Y}_{1.} - \bar{Y}_{..} \\ \bar{Y}_{2.} - \bar{Y}_{..} \\ \bar{Y}_{3.} - \bar{Y}_{..} \end{bmatrix}$$

Example: One-way ANOVA

- Solution A = Cell Means Model
- Solution B = Sets Baseline Constraint for Group 3
- Solution C = Sets Baseline Constraint for Group 1
- Solution D = Sets Sum to Zero Constraint

Evaluating Generalized Inverses

Several algorithms for getting generalized inverses, for example,

Algorithm 1:

1. Find any $r \times r$ nonsingular submatrix of A where r=rank(A). Call this matrix W.
2. Invert and transpose W, i.e., compute $(W^{-1})^T$.
3. Replace each element of W in A with the corresponding element of $(W^{-1})^T$
4. Replace all other elements in A with zeros.
5. Transpose the resulting matrix to obtain G, a generalized inverse for A.

Projection Matrix

Define the projection matrix P_X to be

$$P_X = X(X^T X)^- X^T$$

where $(X^T X)^-$ is a generalized inverse matrix for $X^T X$.

- If X is full rank, the generalized inverse matrix is the usual inverse matrix: $(X^T X)^{-1}$.
- P_X is an orthogonal projection operator onto the column space of X (the set of all possible linear combinations of the columns of X).

Properties of P_X

- P_X is symmetric ($P_X = P_X^T$)
- $P_X X = X$
- P_X is idempotent ($P_X P_X = P_X$)

$$P_X P_X = P_X X(X^T X)^- X^T = X(X^T X)^- X^T = P_X$$

- $P_X \mathbf{u} = \mathbf{u}$ for any vector $\mathbf{u}$ in the space spanned by the columns of X
- $\text{rank}(X) = rank(P_X) = tr(P_X)$
- $P_X = X(X^T X)^- X^T$ is the same matrix for all generalized inverses $(X^T X)^-$ of $X^T X$.

Uniqueness of Mean Estimation

The estimation of mean vector (predicted response vector)

$$\hat{\mathbf{Y}} = X\mathbf{b} = X(X^TX)^{-}X^T\mathbf{Y} = P_X\mathbf{Y}$$

is unique.

- $\hat{\mathbf{Y}} = P_X\mathbf{Y}$ is invariant to the choice of $(X^TX)^{-}$.
- For any solution $\mathbf{b} = (X^TX)^{-}X^T\mathbf{Y}$ to the normal equations, $\hat{\mathbf{Y}} = X\mathbf{b} = P_X\mathbf{Y}$.

Example: One-Way ANOVA

<u>Solution A:</u> Effects Model

$$\hat{\mathbf{Y}} = X\mathbf{b} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \end{bmatrix}$$

Example: One-Way ANOVA

Solution B: Effects Model

$$\hat{\mathbf{Y}} = X\mathbf{b} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{Y}_{3.} \\ \bar{Y}_{1.} - \bar{Y}_{3.} \\ \bar{Y}_{2.} - \bar{Y}_{3.} \\ 0 \end{bmatrix} = \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \end{bmatrix}$$

Example: One-Way ANOVA

Solution C: Effects Model

$$\hat{\mathbf{Y}} = X\mathbf{b} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{Y}_{1.} \\ 0 \\ \bar{Y}_{2.} - \bar{Y}_{1.} \\ \bar{Y}_{3.} - \bar{Y}_{1.} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \end{bmatrix}$$

Example: One-Way ANOVA

Solution D: Effects Model

$$\hat{\mathbf{Y}} = X\mathbf{b} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{Y}_{..} \\ \bar{Y}_{1.} - \bar{Y}_{..} \\ \bar{Y}_{2.} - \bar{Y}_{..} \\ \bar{Y}_{3.} - \bar{Y}_{..} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \end{bmatrix}$$

Residuals

The vector of residuals is

$$\begin{aligned}
\mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}} \\
&= \mathbf{Y} - X\mathbf{b} \\
&= \mathbf{Y} - P_X\mathbf{Y} \\
&= (I - P_X)\mathbf{Y}
\end{aligned}$$

- $I - P_X$ is also a projection matrix and it projects $\mathbf{Y}$ onto the space orthogonal to the space spanned by the columns of X.
- Since the OLS Estimator $\mathbf{b}$ minimizes the function

$$(\mathbf{Y} - X\mathbf{b})^T(\mathbf{Y} - X\mathbf{b})$$

it minimizes the function $\mathbf{e}^T\mathbf{e}$

Properties of $I - P_X$

- $I - P_X$ is symmetric
- $I - P_X$ is idempodent

$$(I - P_X)(I - P_X) = I - P_X - P_X + P_X P_X = I - P_X - P_X + P_X = I - P_X$$

- $(I - P_X)P_X = P_X - P_X P_X = P_X - P_X = 0$
- $(I - P_X)X = X - P_X X = X - X = 0$
- Partition X as $X = [X_1|X_2|\cdots|X_k]$ then $(I - P_X)X_j = 0$
- $(I - P_X)\mathbf{u} = \mathbf{0}$ for any vector $\mathbf{u}$ in the space spanned by the columns of X

Uniqueness of Residuals

Because the projection operator $P_X = X(X^TX)^-X^T$ is invariant with respect to the choice of $(X^TX)^-$, the residuals are invariant with respect to the choice of $(X^TX)^-$, that is,

$$\mathbf{e} = \mathbf{Y} - X\mathbf{b} = (I - P_X)\mathbf{Y}$$

is the same for any solution

$$\mathbf{b} = (X^TX)^-X^T\mathbf{Y}$$

to the normal equations.

Unit 3

Linear Model Theory:

Estimability

Identifiable

For a linear model $E(\mathbf{Y}) = X\beta$, the parameter vector β is *identifiable* if $X\beta_1 = X\beta_2$ implies $\beta_1 = \beta_2$

Identifiability and Estimability

- Only *identifiable* parameters can be estimated
- Linear functions of identifiable parameters are called *estimable*
- Unbiased estimators can be found for estimable functions of model parameters

Example: Identifiability

One-Way ANOVA Cell Means Model

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \end{bmatrix}$$

Example: Identifiability

Let $\boldsymbol{\beta}_1 = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 \end{pmatrix}^T$

$$\begin{bmatrix} E(Y_{11}) \\ E(Y_{12}) \\ E(Y_{13}) \\ E(Y_{14}) \\ E(Y_{21}) \\ E(Y_{22}) \\ E(Y_{23}) \\ E(Y_{24}) \\ E(Y_{31}) \\ E(Y_{32}) \\ E(Y_{33}) \\ E(Y_{34}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_1 \\ \beta_1 \\ \beta_1 \\ \beta_2 \\ \beta_2 \\ \beta_2 \\ \beta_2 \\ \beta_3 \\ \beta_3 \\ \beta_3 \\ \beta_3 \end{bmatrix}$$

Example: Identifiability

Let $\boldsymbol{\beta}_2 = \begin{pmatrix} \beta_1^* & \beta_2^* & \beta_3^* \end{pmatrix}^T$

$$\begin{bmatrix} E(Y_{11}) \\ E(Y_{12}) \\ E(Y_{13}) \\ E(Y_{14}) \\ E(Y_{21}) \\ E(Y_{22}) \\ E(Y_{23}) \\ E(Y_{24}) \\ E(Y_{31}) \\ E(Y_{32}) \\ E(Y_{33}) \\ E(Y_{34}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \beta_3^* \end{bmatrix} = \begin{bmatrix} \beta_1^* \\ \beta_1^* \\ \beta_1^* \\ \beta_1^* \\ \beta_2^* \\ \beta_2^* \\ \beta_2^* \\ \beta_2^* \\ \beta_3^* \\ \beta_3^* \\ \beta_3^* \\ \beta_3^* \end{bmatrix}$$

Example: Identifiability

One-Way ANOVA Cell Means Model

- For $X\beta_1 = X\beta_2$, we must have $\beta_1 = \beta_2$
- For this model, β is identifiable.
- The vector of response means uniquely determines the values of the parameter vector β.

Example: Identifiability

One-Way ANOVA Effects Model

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \end{bmatrix}$$

Example: Identifiability

Let $\beta_1 = \begin{pmatrix} \mu_3 & \mu_1 - \mu_3 & \mu_2 - \mu_3 & 0 \end{pmatrix}^T$

$$\begin{bmatrix} E(Y_{11}) \\ E(Y_{12}) \\ E(Y_{13}) \\ E(Y_{14}) \\ E(Y_{21}) \\ E(Y_{22}) \\ E(Y_{23}) \\ E(Y_{24}) \\ E(Y_{31}) \\ E(Y_{32}) \\ E(Y_{33}) \\ E(Y_{34}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_3 \\ \mu_1 - \mu_3 \\ \mu_2 - \mu_3 \\ 0 \end{bmatrix}$$

Example: Identifiability

Let $\beta_2 = \begin{pmatrix} 0 & \mu_1 & \mu_2 & \mu_3 \end{pmatrix}^T$

$$\begin{bmatrix} E(Y_{11}) \\ E(Y_{12}) \\ E(Y_{13}) \\ E(Y_{14}) \\ E(Y_{21}) \\ E(Y_{22}) \\ E(Y_{23}) \\ E(Y_{24}) \\ E(Y_{31}) \\ E(Y_{32}) \\ E(Y_{33}) \\ E(Y_{34}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

Example: Identifiability

One-Way ANOVA Effects Model

- $X\beta_1 = X\beta_2$ but $\beta_1 \neq \beta_2$, so β is not identifiable.
- The vector of response means does not uniquely determine the values of the parameter vector β.

Estimable Functions

An *estimable function* is a linear function of identifiable parameters

- Estimable functions are reasonable things to estimate
- Estimable functions have the same interpretation regardless of the constraints placed on the parameters to get a solution to the normal equations
- Least squares estimates of estimable functions are not affected by the choice of constraints placed on the parameters to get a solution to the normal equations

Estimable Functions

For a linear model $\mathbf{Y} = X\beta + \epsilon$, we say that a linear function of β, $C\beta$, is *estimable* if

$$C\beta = AX\beta = AE(\mathbf{Y})$$

for some matrix A.

Note that:

- C is a $m \times k$ matrix of constants that defines m estimable linear functions of the parameters, and $C = AX$ for some A

Example: Estimable Functions

One Way ANOVA Effects model

$$\begin{bmatrix} E(Y_{11}) \\ E(Y_{12}) \\ E(Y_{13}) \\ E(Y_{14}) \\ E(Y_{21}) \\ E(Y_{22}) \\ E(Y_{23}) \\ E(Y_{24}) \\ E(Y_{31}) \\ E(Y_{32}) \\ E(Y_{33}) \\ E(Y_{34}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} \mu + \alpha_1 \\ \mu + \alpha_1 \\ \mu + \alpha_1 \\ \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_2 \\ \mu + \alpha_2 \\ \mu + \alpha_2 \\ \mu + \alpha_3 \\ \mu + \alpha_3 \\ \mu + \alpha_3 \\ \mu + \alpha_3 \end{bmatrix}$$

Example: Estimable Functions

Show that $\mu + \alpha_1 = [1\ 1\ 0\ 0]\beta$ is estimable.

Let $A = \left[\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\right]$

$$\begin{aligned} AX\beta = AE(\mathbf{Y}) &= \frac{1}{4}E(Y_{11}) + \frac{1}{4}E(Y_{12}) + \frac{1}{4}E(Y_{13}) + \frac{1}{4}E(Y_{14}) \\ &= \frac{1}{4}(\mu + \alpha_1) + \frac{1}{4}(\mu + \alpha_1) + \frac{1}{4}(\mu + \alpha_1) + \frac{1}{4}(\mu + \alpha_1) \\ &= \mu + \alpha_1 \end{aligned}$$

Alternatively, let $A = [1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$

$$AX\beta = AE(\mathbf{Y}) = E(Y_{11}) = \mu + \alpha_1$$

Example: Estimable Functions

Show that $\begin{bmatrix} \mu + \alpha_2 \\ \mu + \alpha_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \beta$ is estimable.

Let $A = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

$$AX\beta = AE(\mathbf{Y}) = \begin{bmatrix} E(Y_{21}) \\ E(Y_{31}) \end{bmatrix} = \begin{bmatrix} \mu + \alpha_2 \\ \mu + \alpha_3 \end{bmatrix}$$

Example: Estimable Functions

Show that $\alpha_1 - \alpha_2 = [0\ 1\ \ -1\ 0]\beta$ is estimable

$$\begin{aligned} \alpha_1 - \alpha_2 &= (\mu + \alpha_1) - (\mu + \alpha_2) \\ &= E(Y_{11}) - E(Y_{21}) \\ &= [1\ 0\ 0\ 0\ \text{-}1\ 0\ 0\ 0\ 0\ 0\ 0\ 0]E(\mathbf{Y}) \\ &= AE(\mathbf{Y}) \end{aligned}$$

Example: Estimable Functions

Show that $\begin{bmatrix} \alpha_2 - \alpha_3 \\ 2\mu + 3\alpha_1 - \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 2 & 3 & -1 & 0 \end{bmatrix} \beta$ is estimable

$$\begin{aligned} \begin{bmatrix} \alpha_2 - \alpha_3 \\ 2\mu + 3\alpha_1 - \alpha_2 \end{bmatrix} &= \begin{bmatrix} (\mu + \alpha_2) - (\mu + \alpha_3) \\ 3(\mu + \alpha_1) - (\mu + \alpha_2) \end{bmatrix} \\ &= \begin{bmatrix} E(Y_{21}) - E(Y_{31}) \\ 3E(Y_{11}) - E(Y_{21}) \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu+\alpha_1 \\ \mu+\alpha_1 \\ \mu+\alpha_1 \\ \mu+\alpha_1 \\ \mu+\alpha_2 \\ \mu+\alpha_2 \\ \mu+\alpha_2 \\ \mu+\alpha_2 \\ \mu+\alpha_3 \\ \mu+\alpha_3 \\ \mu+\alpha_3 \\ \mu+\alpha_3 \end{bmatrix}$$

$$= AE(\mathbf{Y})$$

Functions that are Not Estimable

Examples:

$$\mu,\ \alpha_1,\ \alpha_2,\ \alpha_3,\ 3\alpha_1,\ \alpha_1 + \alpha_2$$

To show that

$$c^T\boldsymbol{\beta} = c_1\mu + c_2\alpha_1 + c_3\alpha_2 + c_4\alpha_3$$

is <u>not</u> estimable, one must show that there is no non-random matrix A for which

$$c^T\boldsymbol{\beta} = c_1\mu + c_2\alpha_1 + c_3\alpha_2 + c_4\alpha_3 = AX\boldsymbol{\beta} = AE(\mathbf{Y})$$

Example: Non-Estimable Functions:

Show $\alpha_1 = [0\ 1\ 0\ 0]\beta$ is not estimable.

For α_1 to be estimable, we would need to find a matrix A such that

$$\begin{aligned}
\alpha_1 &= AE(\mathbf{Y}) \\
&= a_1E(Y_{11}) + a_2E(Y_{12}) + a_3E(Y_{13}) + a_4E(Y_{14}) \\
&\quad + a_5E(Y_{21}) + a_6E(Y_{22}) + a_7E(Y_{23}) + a_8E(Y_{24}) \\
&\quad + a_9E(Y_{31}) + a_{10}E(Y_{32}) + a_{11}E(Y_{33}) + a_{12}E(Y_{34}) \\
&= (a_1 + a_2 + a_3 + a_4)(\mu + \alpha_1)
\end{aligned}$$

Non-Estimable Functions: Example

This implies that

$$0 = (a_5 + a_6 + a_7 + a_8) = (a_9 + a_{10} + a_{11} + a_{12})$$

and

$$\begin{aligned}\alpha_1 &= (a_1 + a_2 + a_3 + a_4)(\mu + \alpha_1) \\ &= (a_1 + a_2 + a_3 + a_4)\mu + (a_1 + a_2 + a_3 + a_4)\alpha_1\end{aligned}$$

This is not possible, so α_1 is not estimable.

Rules for Estimable Functions

For a linear model $\mathbf{Y} = X\boldsymbol{\beta} + \epsilon$

- The expectation of any observation is estimable.
- A linear combination of estimable functions is estimable.
- Each element of $\boldsymbol{\beta}$ is estimable if and only if $\text{rank}(X) = k =$ number of columns in X.
- Every $\mathbf{c}^T\boldsymbol{\beta}$ is estimable if and only if $\text{rank}(X) = k =$ number of columns in X.
- Let X_j be the j^{th} column of X. β_j is not estimable if and only if $X_j = \sum_{j \neq l} c_l X_l$ for some set of scalars $\{\ c_i :\ j \neq l\}$.

Estimable Functions: Example

- Multiple Linear Regression - show β_j is estimable for all j

  For a multiple linear regression model, the design matrix X is typically full rank (no perfect correlation among predictors). So every element of $\boldsymbol{\beta}$ is estimable.

- Multiple Linear Regression - show $\mu_{Y|x_1,x_2,\ldots,x_k}$ is estimable

  Since every element of $\boldsymbol{\beta}$ is estimable, a linear combination of $\boldsymbol{\beta}$ is also estimable.

$$\mu_{Y|x_1,x_2,\ldots,x_k} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

Non-Estimable Functions: Example

- One-Way ANOVA Effects Model - show μ is not estimable.

  Let X_j be the j^{th} column in the design matrix X.
  $\mu = \beta_1$ is not estimable since we can write

$$X_1 = X_2 + X_3 + X_4 = 1 * X_2 + 1 * X_3 + 1 * X_4$$

Estimable Functions

For a linear model

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

- Definitions of estimable functions of the elements of the parameter vector $\boldsymbol{\beta}$ depend on the linear model for the expected responses $E(\mathbf{Y}) = X\boldsymbol{\beta}$.
- No assumption is made about $Var(\mathbf{Y})$ or $Var(\boldsymbol{\epsilon})$ or the shape of the distribution of $\mathbf{Y}$ or $\boldsymbol{\epsilon}$.

Least Squares Estimator for Estimable Functions

Let $\mathbf{b} = (X^T X)^{-} X^T \mathbf{Y}$ be a solution to the normal equations. For any estimable function $C\beta$, the least squares estimator for this estimable function is $C\mathbf{b}$ and is unique.

This means that $C\mathbf{b}$ has the same value regardless of

- *constraints placed on parameters*
- *the choice of the generalized inverse matrix*

The Gauss-Markov Theorem

For the Gauss-Markov model, $\mathbf{Y} = X\beta + \epsilon$, with

$$E(\mathbf{Y}) = X\beta \text{ and } Var(\mathbf{Y}) = \sigma^2 I,$$

the OLS estimator $C\mathbf{b}$ of an estimable function $C\beta$ is the unique best linear unbiased estimator (BLUE) of this estimable function.

'Best' means out of all possible linear unbiased estimators, the one with the smallest variance.

Questions?

Contact me:

Email: dmommen@iastate.edu
Student Office Hours: Thursdays @ 10-11 am