STAT 500

Estimation in Linear Models

Ordinary Least Squares (OLS) Estimator

For a linear model with $E(\mathbf{Y}) = X\boldsymbol{\beta}$, 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 $\boldsymbol{\beta}$.

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 $\boldsymbol{\beta}$ 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.

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 $\mathbf{b}$ 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 $\mathbf{b}$ is unique.

OLS Estimator – One-way ANOVA

Cell Means Model

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

OLS Estimator – One-way ANOVA

X is full rank: rank$(X) = 3$

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

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

OLS Estimator

If 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.

One-Way ANOVA Model

Effects model

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

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

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

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

OLS Estimation: One-Way ANOVA

Solution A: A generalized inverse for X^TX is

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

and a solution to the normal equations is

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

OLS Estimation: One-Way ANOVA

Solution B: Another generalized inverse for X^TX is

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

and a solution to the normal equations is

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

OLS Estimation: One-Way ANOVA

Solution C: Another generalized inverse for $X^T X$ is

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

and a solution to the normal equations is

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

OLS Estimation: 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

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

OLS Estimation: 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:

(i) Find any $r \times r$ nonsingular submatrix of A where r=rank(A). Call this matrix W.

(ii) Invert and transpose W, ie., compute $(W^{-1})^T$.

(iii) Replace each element of W in A with the corresponding element of $(W^{-1})^T$

(iv) Replace all other elements in A with zeros.

(v) 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

- $P_X X = X$

- 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

- 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, $\hat{\mathbf{Y}} = X\mathbf{b} = P_X\mathbf{Y}$.

$\hat{\mathbf{Y}}$: One-Way ANOVA

Solution A: Effects Model

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

$\hat{\mathbf{Y}}$: One-Way ANOVA

Solution B: Effects Model

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

$\hat{\mathbf{Y}}$: One-Way ANOVA

Solution C: Effects Model

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

$\hat{\mathrm{Y}}$: 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

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

$I - P_X$ 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

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

- $(I - P_X)P_X = P_X - P_XP_X = P_X - P_X = 0$
- $(I - P_X)X = X - P_XX = X - X = 0$
- 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 $P_X = X(X^T X)^- X^T$ is invariant with respect to the choice of $(X^T X)^-$, the residuals are invariant with respect to the choice of $(X^T X)^-$, that is,

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

is the same for any solution

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

to the normal equations.