37 Lecture 37: April 26

Last time

- Theoretical background of linear model

Today

- Course evaluation (5/17)
- HW3 review
- Theoretical background of linear models cont.
  - Estimability
  - Idempotent matrix
  - Projections
  - Geometry of least squares solution

Additional reference

Course notes by Dr. Hua Zhou
"A Primer on Linear Models" by Dr. John F. Monahan

Estimable function

Assume the linear mean model: $\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathrm{E}(\mathbf{e}) = \mathbf{0}$. One main interest is estimation of the underlying parameter $\mathbf{b}$. Can $\mathbf{b}$ be estimated or what functions of $\mathbf{b}$ can be estimated?

- A parametric function $\mathbf{\Lambda b}$, $\mathbf{\Lambda} \in \mathbb{R}^{m \times p}$ is said to be (linearly) estimable if there exists an affinely unbiased estimator of $\mathbf{\Lambda b}$ for all $\mathbf{b} \in \mathbb{R}^p$. That is there exist constants $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{c} \in \mathbb{R}^m$ such that $\mathrm{E}(\mathbf{A}\mathbf{y} + \mathbf{c}) = \mathbf{\Lambda b}$ for all $\mathbf{b}$.
- Theorem: Assuming the linear mean model, the parametric function $\mathbf{\Lambda b}$ is (linearly) estimable if and only if $\mathcal{C}(\mathbf{\Lambda}) \subset \mathcal{C}(\mathbf{X}^T)$, or equivalently $\mathcal{N}(\mathbf{X}) \subset \mathcal{N}(\mathbf{\Lambda})$.
"$\mathbf{\Lambda b}$ is estimable $\iff$ the row space of $\mathbf{\Lambda}$ is contained in the row space of $\mathbf{X}$ $\iff$ the null space of $\mathbf{X}$ is contained in the null space of $\mathbf{\Lambda}$."
Proof:
- $\lambda^T\mathbf{b}$ is linearly estimable if and only if $\lambda^T\mathbf{b}$ is a linear combination of the components in $\mu_Y = \mathrm{E}(\mathbf{Y})$
- Corollary: $\mathbf{Xb}$ is estimable.
"Expected value of any observation $\mathrm{E}(y_i)$ and their linear combinations are estimable."
- Corollary: If $\mathbf{X}$ has full column rank, then any linear combinations of $\mathbf{b}$ are estimable.

- If $\mathbf{\Lambda b}$ is (linearly) estimable, then its *least squares estimator* $\mathbf{\Lambda}\hat{\mathbf{b}}$ is invariant to the choice of the least squares solution $\hat{\mathbf{b}}$.
  Proof:

- The least squares estimator $\mathbf{\Lambda}\hat{\mathbf{b}}$ is a linearly unbiased estimator of $\mathbf{\Lambda b}$. *Proof:*

Estimability example: One-way ANOVA model

Consider the following example with one-way ANOVA model.

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad i = 1, 2, 3, \ j = 1, 2$$

In matrix form:

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

Note: replication doesn't help with estimability. What functions of $\lambda^T \mathbf{b}$ are estimable?
Solutions:

Idempotent matrix

Assume $\mathbf{A} \in \mathbb{R}^{n \times n}$.

- A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is <u>idempotent</u> if and only if $\mathbf{A}^2 (= \mathbf{AA}) = \mathbf{A}$.
- Any idempotent matrix $\mathbf{A}$ is a generalized inverse of itself.
- The only idempotent matrix of full rank is $\mathbf{I}$.
  Proof. Interpretation: all idempotent matrices are singular except for the identity matrix.
- $\mathbf{A}$ is idempotent if and only if $\mathbf{A}^T$ is idempotent if and only if $\mathbf{I}_n - \mathbf{A}$ is idempotent.
- For a general matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, the matrices $\mathbf{A}^- \mathbf{A}$ and $\mathbf{A}\mathbf{A}^-$ are idempotent and

$$\begin{aligned} \operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^- \mathbf{A}) &= \operatorname{rank}(\mathbf{A}\mathbf{A}^-) \\ \operatorname{rank}(\mathbf{I}_n - \mathbf{A}^- \mathbf{A}) &= n - \operatorname{rank}(\mathbf{A}) \\ \operatorname{rank}(\mathbf{I}_m - \mathbf{A}\mathbf{A}^-) &= m - \operatorname{rank}(\mathbf{A}). \end{aligned}$$

Projection

- A matrix $\mathbf{P} \in \mathbb{R}^{m \times n}$ is a <u>projection</u> onto a vector space $\mathcal{V}$ if and only if
  1. $\mathbf{P}$ is idempotent
  2. $\mathbf{Px} \in \mathcal{V}$ for any $\mathbf{x} \in \mathbb{R}^n$

3. $\mathbf{Pz} = \mathbf{z}$ for any $\mathbf{z} \in \mathcal{V}$.

- Any idempotent matrix $\mathbf{P}$ is a projection onto its own column space $\mathcal{C}(\mathbf{P})$.
  Proof:

- $\mathbf{AA}^-$ is a projection onto the column space $\mathcal{C}(\mathbf{A})$.
  Proof:

- The projection matrix
$$\underset{n\times n}{\mathbf{P_X}} = \underset{n\times p}{\mathbf{X}} \underset{p\times p}{(\mathbf{X}^T\mathbf{X})^-} \underset{p\times n}{\mathbf{X}^T}$$
  is unique.
  Proof:

- Proposition: Let $\mathbf{X}, \mathbf{A}, \mathbf{B}$ be matrices, then $\mathbf{X}^T\mathbf{XA} = \mathbf{X}^T\mathbf{XB}$ if and only if $\mathbf{XA} = \mathbf{XB}$.
  Proof:

- $\underset{n\times n}{\mathbf{P_X}} \underset{n\times p}{\mathbf{X}} = \underset{n\times p}{\mathbf{X}}$
  Proof:

- Predicted values $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{b}}_{ls}$ are invariant to choice of solution to the normal equation, where
$$\hat{\mathbf{b}}_{ls} = (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{Y}$$
  is not necessarily unique.
  Proof:

- Start with $\mathbf{P_X X} = \mathbf{X}$, we have $\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X} = \mathbf{X}$. Therefore, $(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ is a generalized inverse of $\mathbf{X}$ which is sometimes called the least-squares inverse. And $\mathbf{P_X}$ is a projection onto $\mathcal{C}(\mathbf{X})$.

Geometry of least squares

- $\mathbf{P}_\mathbf{X}^2 = \mathbf{P_X}$ and $\hat{\mathbf{Y}} = \mathbf{P_X Y}$ is unique.

- Recall the column space of $\mathbf{X}$ is $\mathcal{C}(\mathbf{X}) = \left\{ \underset{n\times 1}{\mathbf{y}} : \mathbf{y} = \mathbf{X} \underset{p\times 1}{\mathbf{b}} \text{ for some } \mathbf{b} \right\}$

- The vector in $\mathcal{C}(\mathbf{X})$ that is closest in terms of squared norm (L_2 norm: $||\mathbf{a} - \mathbf{b}||_2 = \sqrt{(\mathbf{a} - \mathbf{b})^T(\mathbf{a} - \mathbf{b})}$) to $\mathbf{Y}$ is given by $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{b}}_{ls} = \mathbf{P_X Y}$.
  Proof:

- $\hat{\mathbf{Y}} \in \mathcal{C}(\mathbf{X})$

- $\underset{n\times 1}{\hat{\mathbf{e}}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{P_X})\mathbf{Y} \in \mathcal{N}(\mathbf{X}^T)$ where $\mathcal{N}(\mathbf{X}^T) = \left\{ \underset{n\times 1}{\mathbf{v}} : \mathbf{X}^T\mathbf{v} = \mathbf{0} \right\}$ is the null space of $\mathbf{X}^T$.
  Proof: