# 35 Lecture 35: April 21

## Last time

- nested design
- Two-factor designs

## Today

- HW3 deadline extended to Friday 04/23 midnight.
- Theoretical background of linear models

### Additional reference

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

## Linear Models in the matrix form

Recall the matrix form of the linear model

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}} \underset{p\times 1}{\beta} + \underset{n\times 1}{\epsilon}$$

### Simple linear regression model

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

### Multiple linear regression model

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1,p-1} \\ 1 & x_{21} & \dots & x_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{n,p-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

One-way ANOVA model

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1,n_1} \\ y_{21} \\ \vdots \\ y_{2,n_2} \\ \vdots \\ y_{a,1} \\ \vdots \\ y_{a,n_a} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & & & \\ \mathbf{1}_{n_2} & & \mathbf{1}_{n_2} & & \\ \vdots & & & \ddots & \\ \mathbf{1}_{n_a} & & & & \mathbf{1}_{n_a} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1,n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2,n_2} \\ \vdots \\ \epsilon_{a,1} \\ \vdots \\ \epsilon_{a,n_a} \end{bmatrix}$$

Two-way ANOVA model without interaction  Model $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$, $i = 1, \ldots, a$ ($a$ levels in factor 1), $j = 1, \ldots, b$ (b levels in factor 2), and $k = 1, \ldots, n_{ij}$ ($n_{ij}$ observations in the $(i, j)$-th cell). In total we have $n = \sum_{i,j} n_{ij}$ observations and $p = a + b + 1$ parameters. For simplicity, we consider the case without replicates, i.e., $n_{ij} = 1$ and only write out $\mathbf{X}\beta$. Note adding more replicates to each cell does *not* change the rank of $\mathbf{X}$.

$$\mathrm{E}(\mathbf{y}) = \mathbf{X}\beta = \begin{bmatrix} \mathbf{1}_b & \mathbf{1}_b & & & & \mathbf{I}_b \\ \mathbf{1}_b & & \mathbf{1}_b & & & \mathbf{I}_b \\ \vdots & & & \ddots & & \vdots \\ \mathbf{1}_b & & & & \mathbf{1}_b & \mathbf{I}_b \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \\ \beta_1 \\ \vdots \\ \beta_b \end{bmatrix}$$

Two-way ANOVA with interaction  Model $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$, $i = 1, \ldots, a$ ($a$ levels in factor 1), $j = 1, \ldots, b$ (b levels in factor 2), and $k = 1, \ldots, n_{ij}$ ($n_{ij}$ observations in the $(i, j)$-th cell). In total we have $n = \sum_{i,j} n_{ij}$ observations and $p = 1 + a + b + ab$ parameters. For simplicity, we consider the case without replicates, i.e., $n_{ij} = 1$ and only write out $\mathbf{X}\beta$.

Note adding more replicates to each cell does *not* change the rank of $\mathbf{X}$.

$$\mathrm{E}(\mathbf{y}) = \mathbf{X}\beta = \begin{bmatrix} \mathbf{1}_b & \mathbf{1}_b & & & & \mathbf{I}_b & \mathbf{I}_b & & & \\ \mathbf{1}_b & & \mathbf{1}_b & & & \mathbf{I}_b & & \mathbf{I}_b & & \\ \vdots & & & \ddots & & \vdots & & & \ddots & \\ \mathbf{1}_b & & & & \mathbf{1}_b & \mathbf{I}_b & & & & \mathbf{I}_b \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \\ \beta_1 \\ \vdots \\ \beta_b \\ \gamma_{11} \\ \vdots \\ \vdots \\ \gamma_{ab} \end{bmatrix}$$

For all the above models, we have 5the most general assumption over the error term, i.e. $\epsilon \sim \mathcal{N}(0, \sigma^2\mathbf{I})$.

**Mixed effects models** For mixed effects models, we generally have

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

- $\mathbf{X} \in \mathbb{R}^{n\times p}$ is a design matrix for fixed-effects $\mathbf{b} \in \mathbb{R}^p$
- $\mathbf{Z} \in \mathbb{R}^{n\times q}$ is a design matrix for random-effects $\mathbf{u} \in \mathbb{R}^q$
- The most general assumption is $\mathbf{e} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{R})$, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}_q, \mathbf{G})$, and $\mathbf{e}$ is independent of $\mathbf{u}$.

In many applications, $\mathbf{e} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$ and

$$\mathbf{Z}\mathbf{u} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_m) \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_m \end{pmatrix} = \mathbf{Z}_1\mathbf{u}_1 + \cdots + \mathbf{Z}_m\mathbf{u}_m,$$

where $\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}_{q_i}, \sigma_i^2\mathbf{I}_{q_i})$, $\sum_{i=1}^m q_i = q$. $\mathbf{e}$ and $\mathbf{u}_i$, $i = 1, \ldots, m$, are jointly independent. Then the covariance of responses $\mathbf{y}$

$$\mathbf{V}(\sigma^2, \sigma_1^2, \ldots, \sigma_m^2) = \sigma^2\mathbf{I} + \sum_{i=1}^m \sigma_i^2\mathbf{Z}_i\mathbf{Z}_i^T$$

## Linear equations and generalized inverse

For the linear model

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\,\underset{p\times 1}{\mathbf{b}} + \underset{n\times 1}{\mathbf{e}},$$

we obtain the least square estimator by minimize the objective function $Q(\mathbf{b}) = \sum_{i=1}^{n} e_i^2 = (\mathbf{Y} - \mathbf{X}\mathbf{b})^T(\mathbf{Y} - \mathbf{X}\mathbf{b})$. By taking derivative with respect to $\mathbf{b}$ and setting it to zero, we get

$$\left(\frac{\partial Q}{\partial \mathbf{b}}\right)^T = \left(\frac{\partial Q}{\partial b_1}, \frac{\partial Q}{\partial b_2}, \ldots, \frac{\partial Q}{\partial b_p}\right)^T = \left[\frac{\partial\left(\mathbf{Y}^T\mathbf{Y} - 2\mathbf{Y}^T\mathbf{X}\mathbf{b} + \mathbf{b}^T\mathbf{X}^T\mathbf{X}\mathbf{b}\right)}{\partial \mathbf{b}}\right]^T = -2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\mathbf{b}$$

where we used the fact that for constant vector $\mathbf{a} \in \mathbb{R}^{p\times 1}$, constant matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$ and $\mathbf{x} \in \mathbb{R}^{p\times 1}$, we have the two derivatives:

1. $\frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^T$
2. $\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{x}^T(\mathbf{A} + \mathbf{A}^T)$

By setting $\left(\frac{\partial Q}{\partial \mathbf{b}}\right)^T = \mathbf{0}_{p\times 1}$, we get the Normal equations

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

## Consistency

Assume $\mathbf{A} \in \mathbb{R}^{m\times n}$
*Definition:* The linear system $\mathbf{A}\mathbf{x} = c$ is consistent if there exists an $\mathbf{x}^*$ such that $\mathbf{A}\mathbf{x}^* = \mathbf{c}$.

- If $\mathbf{A}$ is square and $\mathbf{A}^{-1}$ exists, then $\mathbf{x} = \mathbf{A}^{-1}\mathbf{c}$.
- Proposition (g1): If $\mathbf{A}\mathbf{x} = \mathbf{c}$ is consistent, and if $\mathbf{G}$ is any matrix such that $\underset{m\times n}{\mathbf{A}}\ \underset{n\times m}{\mathbf{G}}\ \underset{m\times n}{\mathbf{A}} = \underset{m\times n}{\mathbf{A}}$, then $\mathbf{x}^{\psi} = \mathbf{G}\mathbf{c}$ is a solution to $\mathbf{A}\mathbf{x} = \mathbf{c}$.
  *Proof:*
- A matrix $\mathbf{G}$ satisfying $\mathbf{A}\mathbf{G}\mathbf{A} = \mathbf{A}$ is a generalized inverse of $\mathbf{A}$ with notation $\mathbf{A}^-$.
- If $\mathbf{A}$ is square and $\mathbf{A}^{-1}$ exists, then $\mathbf{A}^- = \mathbf{A}^{-1}$ is unique.

## The set of all solutions to $\mathbf{A}\mathbf{x} = \mathbf{c}$

Suppose that $\mathbf{A}\mathbf{x} = \mathbf{c}$ is consistent. Then $\mathbf{x}^*$ is a solution to $\mathbf{A}\mathbf{x} = \mathbf{c}$ if and only if $\mathbf{x}^* = \mathbf{A}^-\mathbf{c} + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{z}$ for some $\mathbf{z}$ and $\mathbf{A}^-$.
*Proof:*

## Moore-Penrose inverse

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

- The Moore-Penrose inverse of $\mathbf{A}$ is a matrix $\mathbf{A}^+ \in \mathbb{R}^{n\times m}$ with the following properties
  1. $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$ (Generalized inverse, $g_1$ inverse, or inner pseudo-inverse)
  2. $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$. (outer pseudo-inverse. Any $g_1$ inverse that satisfies this condition is called a $g_2$ inverse, or reflexive generalized inverse)

3. $\mathbf{A}^+\mathbf{A}$ is symmetric
4. $\mathbf{A}\mathbf{A}^+$ is symmetric

- $\mathbf{A}^+$ exists and is unique for any matrix $\mathbf{A}$.
- In practice, the Moore-Penrose inverse $\mathbf{A}^+$ is easily computed from the singular value decomposition of $\mathbf{A}$.
- $(\mathbf{A}^-)^T$ is a generalized inverse of $\mathbf{A}^T$

### General form of the least squares solution

Now we have derived the general form of the least squares solution with generalized inverse.

$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y} + [\mathbf{I}_p - (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}]\mathbf{q}$$

where $\mathbf{q} \in \mathbb{R}^p$ is arbitrary.

## Positive (semi)definite matrix

Assume $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric (i.e. $\mathbf{A} = \mathbf{A}^T$)

- A real symmetrix matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is positive semi-definite (or nonnegative definite, or p.s.d.) if $\mathbf{x}^T\mathbf{A}\mathbf{x} \geqslant 0$ for all $\mathbf{x}$. Notation $\mathbf{A} \succeq_{p.s.d.} \mathbf{0}$
- E.g., the Gramian matrix $\mathbf{X}^T\mathbf{X}$ is p.s.d.
- We write $\mathbf{A} \succeq_{p.s.d.} \mathbf{B}$ means $\mathbf{A} - \mathbf{B} \succeq_{p.s.d.} \mathbf{0}$
- Cholesky decomposition. Each positive semidefinite matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ can be factorized as $\mathbf{A} = \mathbf{L}\mathbf{L}^T$ for some lower triangular matrix $\mathbf{L} \in \mathbb{R}^{n\times n}$ with nonnegative diagonal entries.
- $\mathbf{A} \in \mathbb{R}^{n\times n}$ is positive semidefinite if and only if $\mathbf{A}$ is a covariance matrix of a random vector.
  *Proof:*

## 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 $\mathbf{\Lambda}\mathbf{b}$, $\mathbf{\Lambda} \in \mathbb{R}^{m\times p}$ is said to be (linearly) estimable if there exists an affinely unbiased estimator of $\mathbf{\Lambda}\mathbf{b}$ 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 $\mathrm{E}(\mathbf{A}\mathbf{y} + \mathbf{c}) = \mathbf{\Lambda}\mathbf{b}$ for all $\mathbf{b}$.
- Theorem: Assuming the linear mean model, the parametric function $\mathbf{\Lambda}\mathbf{b}$ is (linearly) estimable if and only if $\mathcal{C}(\mathbf{\Lambda}) \subset \mathcal{C}(\mathbf{X}^T)$, or equivalently $\mathcal{N}(\mathbf{X}) \subset \mathcal{N}(\mathbf{\Lambda})$.
  "$\mathbf{\Lambda}\mathbf{b}$ 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:*

- 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:*