Advanced Methods in Biostatistics I

Lecture 7

Martin Lindquist

September 19, 2017

Least squares

- Recall the least squares criteria:

$$f(\beta) \quad = \quad ||\mathbf{y} - \mathbf{X}\beta||^2$$

- The solution is obtained by solving the so-called normal equations:

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}.$$

Solution

- Note that the matrix $\mathbf{X}'\mathbf{X}$ retains the same rank as $\mathbf{X}$.
- Hence, if the $n \times p$ design matrix $\mathbf{X}$ has rank p, then $\mathbf{X}'\mathbf{X}$ is a full rank $p \times p$ matrix and invertible, and we can solve the normal equations as:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Less Than Full Rank Model

- However, if the $n \times p$ design matrix $\mathbf{X}$ has rank $r < p$, this does not hold.
- Though θ is uniquely defined as $\theta = \mathbf{X}\beta$, β is not, since the columns of $\mathbf{X}$ are linearly dependent.
- In this setting we say that β is non-identifiable and the normal equations do not have a unique solution for β.

Less Than Full Rank Model

We have three ways to find *a* solution $\hat{\beta}$ and *the* orthogonal projection $\hat{\mathbf{y}}$:

1. Reducing the model to one of full rank.
2. Finding a generalized inverse $(\mathbf{X}'\mathbf{X})^{-}$.
3. Imposing identifiability constraints.

Reducing the Model to One of Full Rank

- Let $\mathbf{X}_1$ consist of r linearly independent columns from $\mathbf{X}$ and let $\mathbf{X}_2$ consist of the remaining columns.
- Then $\mathbf{X_2} = \mathbf{X_1 F}$ because the columns of $\mathbf{X}_2$ are linearly dependent on the columns of $\mathbf{X}_1$.
- Hence,

$$\mathbf{X} = [\mathbf{X_1} \ \mathbf{X_2}] = [\mathbf{X_1} \ \mathbf{X_1 F}] = \mathbf{X_1}[\mathbf{I_{r \times r}} \ \mathbf{F}].$$

Reducing the Model to One of Full Rank

- This is a special case of the rank factorization $\mathbf{X} = \mathbf{KL}$, where $\text{rank}(\mathbf{K}_{n \times r}) = r$ and $\text{rank}(\mathbf{L}_{r \times p}) = r$.
- Now, we can write:

$$\mathbf{X}\beta = \mathbf{KL}\beta = \mathbf{K}\alpha.$$

Reducing the Model to One of Full Rank

- Since $\mathbf{K}$ has full rank, the least squares estimate of α is $\hat{\alpha} = (\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{Y}$.
- The orthogonal projection is:

$$\hat{\mathbf{Y}} = \mathbf{K}\hat{\alpha} = \mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{Y}.$$

- Therefore,

$$\mathbf{P} = \mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'$$

or

$$\mathbf{P} = \mathbf{X_1}(\mathbf{X_1}'\mathbf{X_1})^{-1}\mathbf{X_1}'$$

.

One-way ANOVA with 2 groups

Consider the model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad \text{for } j = 1, \ldots n_i; \quad i = 1, 2$$

or

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix}$$

One-way ANOVA with 2 groups

- Let $\mathbf{X}_1$ consist of the first 2 columns of $\mathbf{X}$.
- Then

$$\mathbf{X} = \mathbf{X}_1 \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix},$$

  and

$$\mathbf{X}\beta = \mathbf{X}_1 \boldsymbol{\alpha},$$

  where

$$\boldsymbol{\alpha} = \begin{pmatrix} \mu + \alpha_2 \\ \alpha_1 - \alpha_2 \end{pmatrix}.$$

One-way ANOVA with 2 groups

- Then

$$\begin{aligned}
\hat{\alpha} &= \begin{pmatrix} n_1 + n_2 & n_1 \\ n_1 & n_1 \end{pmatrix}^{-1} \begin{pmatrix} \sum_j Y_{1j} + \sum_j Y_{2j} \\ \sum_j Y_{1j} \end{pmatrix} \\
&= \begin{pmatrix} n_2^{-1} & -n_2^{-1} \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} \end{pmatrix} \begin{pmatrix} \sum_j Y_{1j} + \sum_j Y_{2j} \\ \sum_j Y_{1j} \end{pmatrix} \\
&= \begin{pmatrix} \bar{Y}_2 \\ \bar{Y}_1 - \bar{Y}_2 \end{pmatrix},
\end{aligned}$$

- Hence, $\hat{\mathbf{Y}} = \mathbf{X}_1 \hat{\alpha} = (\bar{Y}_{1\cdot}, \dots, \bar{Y}_{1\cdot}, \bar{Y}_{2\cdot}, \dots, \bar{Y}_{2\cdot})'$.

Generalized Inverse

Definition

For an $m \times n$ matrix $\mathbf{A}$, a generalized inverse of $\mathbf{A}$ is an $n \times m$ matrix $\mathbf{A}^-$ satisfying $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$.

Generalized Inverse

- A generalized inverse always exists but is not unique except when $\mathbf{A}$ is nonsingular, in which case $\mathbf{A}^{-} = \mathbf{A}^{-1}$.
- Sometimes the generalized inverse is referred to as the conditional inverse.

Example

- Let
$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

- Then the following are all generalized inverses:
$$\mathbf{A}_1^- = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \ \mathbf{A}_2^- = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \ \mathbf{A}_3^- = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Generalized Inverse

Properties

If $\mathbf{G}$ is a generalized inverse of $\mathbf{X}'\mathbf{X}$, then:

- $\mathbf{G}'$ is also a generalized inverse of $\mathbf{X}'\mathbf{X}$,
- $\mathbf{XGX}'\mathbf{X} = \mathbf{X}$, i. e. $\mathbf{GX}'$ is a generalized inverse of $\mathbf{X}$,
- $\mathbf{XGX}'$ is invariant to $\mathbf{G}$,
- $\mathbf{XGX}'$ is symmetric, whether $\mathbf{G}$ is or not.

Finding a Generalized Inverse

Theorem

Let the matrix $\mathbf{W}_{p \times p}$ have rank r and be partitioned as

$$\mathbf{W} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix},$$

where $\mathbf{A}$ has rank r. Then a generalized inverse of $\mathbf{W}$ is

$$\mathbf{W}^{-} = \begin{pmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Finding a Generalized Inverse

- Let $\mathbf{X} = [\mathbf{X_1}, \mathbf{X_2}]$, where $\mathbf{X}_1$ consists of r linearly independent columns from **X.**
- Then a generalized inverse of $\mathbf{X}'\mathbf{X}$ is

$$(\mathbf{X}'\mathbf{X})^{-} = \begin{pmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Solving the normal equations

- A solution to the normal equations is given by

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y} \ .$$

- This can be seen by noting:

$$\begin{aligned} (\mathbf{X}'\mathbf{X})\hat{\beta} &= (\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X})\hat{\beta} \\ &= (\mathbf{X}'\mathbf{X})[(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}] \end{aligned}$$

Solving the normal equations

- Using this result, we can write:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y} = \mathbf{P}\mathbf{Y},$$

  where

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'.$$

- Note that this also gives:

$$\mathbf{P} = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'.$$

Properties

- Recall if $\mathbf{G}$ is a generalized inverse of $\mathbf{X}'\mathbf{X}$, then $\mathbf{X}\mathbf{G}\mathbf{X}'$ is invariant to $\mathbf{G}$.
- Thus, the projection matrix $\mathbf{P}$ is invariant to the choice of $\mathbf{G}$.

One-way ANOVA with 2 groups, continued

- Note we have that

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{pmatrix}.$$

- If $\mathbf{X}_1$ consists of the first 2 columns of $\mathbf{X}$, then

$$(\mathbf{X}_1'\mathbf{X}_1)^{-1} = \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}^{-1} = \begin{pmatrix} n_2^{-1} & -n_2^{-1} \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} \end{pmatrix}.$$

One-way ANOVA with 2 groups, continued

- The generalized inverse of $\mathbf{X}'\mathbf{X}$ is

$$(\mathbf{X}'\mathbf{X})^- = \begin{pmatrix} n_2^{-1} & -n_2^{-1} & 0 \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

One-way ANOVA with 2 groups, continued

- Now a solution to the normal equations is

$$
\begin{aligned}
\hat{\beta} &= \begin{pmatrix} n_2^{-1} & -n_2^{-1} & 0 \\ -n_2^{-1} & n_1^{-1} + n_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \sum_j Y_{1j} + \sum_j Y_{2j} \\ \sum_j Y_{1j} \\ \sum_j Y_{2j} \end{pmatrix} \\
&= \begin{pmatrix} \bar{Y}_2 \\ \bar{Y}_1 - \bar{Y}_2 \\ 0 \end{pmatrix},
\end{aligned}
$$

- Note: $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = (\bar{Y}_{1\cdot}, \ldots, \bar{Y}_{1\cdot}, \bar{Y}_{2\cdot}, \ldots, \bar{Y}_{2\cdot})'$ as before.

Moore-Penrose Inverse

Definition

A matrix $\mathbf{A}^+$ satisfying the following conditions is called the Moore-Penrose inverse:

1. $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$
2. $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$
3. $(\mathbf{A}^+\mathbf{A})' = \mathbf{A}^+\mathbf{A}$
4. $(\mathbf{A}\mathbf{A}^+)' = \mathbf{A}\mathbf{A}^+$

Moore-Penrose Inverse

- $\mathbf{A}^+$ is unique.
- Using the Moore-Penrose inverse provides the minimum norm solution to the least squares problem.
- It is sometimes called the pseudo-inverse.

Singular Value Decomposition

Definition

The singular value decomposition of the $m \times n$ matrix $\mathbf{A}$ is the factorization $\mathbf{A} = \mathbf{USV}^T$, where the $m \times m$ matrix $\mathbf{U}$ and the $n \times n$ matrix $\mathbf{V}$ are orthonormal and the $m \times n$ matrix $\mathbf{S}$ is diagonal with positive real entries organized in descending order.

Moore-Penrose Inverse

- If the singular value decomposition of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{USV}^T$, then the pseudo inverse of $\mathbf{A}$ is $\mathbf{A}^+ = \mathbf{VS}^+\mathbf{U}^T$.
- Here $\mathbf{S}^+$ is obtained by transposing $\mathbf{S}$ and inverting all nonzero entries.

Imposing Identifiability Constraints

- Alternatively, we can impose $s = p - r$ constraints on β to make β uniquely determined (identifiable).
- That is, for any $\boldsymbol{\theta} \in \mathcal{R}(\mathbf{X})$, there is a unique β satisfying

$$\mathbf{X}\beta = \boldsymbol{\theta} \quad \text{and} \quad \mathbf{H}\beta = \mathbf{0}.$$

Imposing Identifiability Constraints

- This can be written

$$\begin{pmatrix} \boldsymbol{\theta} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{H} \end{pmatrix} \boldsymbol{\beta} \equiv \mathbf{G}\boldsymbol{\beta}.$$

- Now when is there a unique solution?

Imposing Identifiability Constraints

Theorem

A unique solution exists if and only if $\mathbf{G}$ has rank p and the rows of $\mathbf{H}$ are linearly independent of the rows of $\mathbf{X}$.

Imposing Identifiability Constraints

Theorem

A unique solution exists if and only if $\mathbf{G}$ has rank p and $\mathbf{H}$ has rank $p - r$.

Imposing Identifiability Constraints

- To estimate β, we solve $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$ and $\mathbf{H}\hat{\beta} = \mathbf{0}$.
- This requires the solution to the augmented normal equations $\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{Y}$ and $\mathbf{H}'\mathbf{H}\hat{\beta} = \mathbf{0}$, or $(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})\hat{\beta} = (\mathbf{G}'\mathbf{G})\hat{\beta} = \mathbf{X}'\mathbf{Y}$.

Imposing Identifiability Constraints

- Therefore,

$$\hat{\beta} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{X}'\mathbf{Y},$$

and

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{P}\mathbf{Y},$$

where

$$\mathbf{P} = \mathbf{X}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{X}'.$$

One-way ANOVA with 2 groups, cont.

- Set $\alpha_1 + \alpha_2 = 0$, i.e.

$$\mathbf{H}\beta \equiv (0, 1, 1)\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = 0.$$

- Suppose $n_1 = n_2 = m$.

One-way ANOVA with 2 groups, cont.

- Then it can be shown that

$$\hat{\beta} = \begin{pmatrix} \bar{Y}_{..} \\ \frac{1}{2}(\bar{Y}_{1.} - \bar{Y}_{2.}) \\ \frac{1}{2}(\bar{Y}_{2.} - \bar{Y}_{1.}) \end{pmatrix}$$

  satisfies the normal equations, and the constraint $\alpha_1 + \alpha_2 = 0$.

- Therefore, we have as before:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = (\bar{Y}_{1.}, \ldots, \bar{Y}_{1.}, \bar{Y}_{2.}, \ldots, \bar{Y}_{2.})'.$$