STAT 714 fa 2023 Lec 02

Estimability, reparameterization, imposing conditions for a unique solution, estimation in a restricted model

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

1. Estimability
2. Reparameterization
3. Imposing conditions for a unique solution
4. Estimation in a restricted model

$X^T X b = X^T y$

e.g. $\underset{\sim}{b} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$ $\quad \underset{\sim}{c}^T \underset{\sim}{b} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^T \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \beta_1$

responses design error term

Throughout, let $\underset{n\times 1}{\mathbf{y}} = \underset{n \times p}{\mathbf{X}}\, \mathbf{b} + \mathbf{e}$, where $\mathbb{E}\mathbf{e} = \mathbf{0}$ (the first assumption we have made).

$p \times 1$, "parameters"

$\underset{\sim}{c}^T \underset{\sim}{b} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^T \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = \beta_1 - \beta_2$

Contrasts in the model parameters

A *contrast* is a linear combination of $\mathbf{b}$, say $\mathbf{c}^T\mathbf{b}$, that we wish to estimate.

Depending on the design $\mathbf{X}$, there may be contrasts the data cannot tell us about.

Estimability of a contrast

A contrast $\mathbf{c}^T\mathbf{b}$ is called *linearly estimable* in the model $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ if there exists a scalar a_0 and a vector $\mathbf{a}$ such that $\mathbb{E}(a_0 + \mathbf{a}^T\mathbf{y}) = \mathbf{c}^T\mathbf{b}$ for all $\mathbf{b}$.

We often drop "linearly" from linearly estimable and just say "estimable."

Which contrasts are estimable?

Result (To know if a contrast is estimable)

A contrast $\mathbf{c}^T\mathbf{b}$ *is estimable in the model* $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ *if and only if* $\mathbf{c} \in \text{Col}\,\mathbf{X}^T$.

See Res 3.1 of Monahan (2008).

Prove the result.

"$\Rightarrow$" $\quad \underset{\sim}{c}^T \underset{\sim}{b}$ is estimable.

Means $\exists$ $a_0 \in \mathbb{R}$ and $\underset{\sim}{a} \in \mathbb{R}^n$ s.t. $\mathbb{E}[a_0 + \underset{\sim}{a}^T \underset{\sim}{y}] = \underset{\sim}{c}^T \underset{\sim}{b}$ $\quad \forall\, \underset{\sim}{b} \in \mathbb{R}^p$

$\Rightarrow \quad a_0 + \underset{\sim}{a}^T \mathbb{E}\underset{\sim}{y} = \underset{\sim}{c}^T \underset{\sim}{b} \quad \forall\, \underset{\sim}{b} \in \mathbb{R}^p$

$\Rightarrow \quad a_0 + \underset{\sim}{a}^T X \underset{\sim}{b} = \underset{\sim}{c}^T \underset{\sim}{b} \quad \forall\, \underset{\sim}{b} \in \mathbb{R}^p$ (try $\underset{\sim}{b} = 0$)

$\Rightarrow \quad a_0 = 0, \quad \underset{\sim}{a}^T X = \underset{\sim}{c}^T \iff \underset{\sim}{c} = X^T \underset{\sim}{a}$

$\Rightarrow \underset{\sim}{c} \in \text{Col}\, X^T.$

"$\Leftarrow$" Let $\underset{\sim}{c} \in \text{Col}\, X^T$.

Then $\exists\, \underset{\sim}{z}$ s.t. $\underset{\sim}{c} = X^T \underset{\sim}{z}$.

Then let $a_0 + \underset{\sim}{a}^T \underset{\sim}{y}$ be an estimator of $\underset{\sim}{c}^T \underset{\sim}{b}$.

Can we find a_0 and $\underset{\sim}{a}$ such that the estimator is unbiased?

We have $\mathbb{E}\left[a_0 + \underset{\sim}{a}^T \underset{\sim}{y}\right] = a_0 + \underset{\sim}{a}^T X \underset{\sim}{b}$.

Can I choose a_0, $\underset{\sim}{a}$ such that

$$a_0 + \underset{\sim}{a}^T X \underset{\sim}{b} = \underset{\sim}{c}^T \underset{\sim}{b} \quad \forall\, \underset{\sim}{b}\ ?$$

Take $a_0 = 0$, and take $\underset{\sim}{a} = \underset{\sim}{z}$

then we have

$$a_0 + \underset{\sim}{a}^T X \underset{\sim}{b} = \underset{\sim}{c}^T \underset{\sim}{b} \quad \forall\, \underset{\sim}{b} \in \mathbb{R}^p.$$

Is $\underset{\sim}{c}^T \underset{\sim}{b}$ estimable? ______

In matrix form. $X\underset{\sim}{b}$

$$X = \begin{bmatrix} \underset{\sim}{1}_{n_1} & \underset{\sim}{1}_{n_1} & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ \underset{\sim}{1}_{n_2} & \underset{\sim}{0} & \underset{\sim}{1}_{n_2} & \cdots & \underset{\sim}{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \underset{\sim}{1}_{n_a} & \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \underset{\sim}{1}_{n_a} \end{bmatrix} \qquad \underset{\sim}{b} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \end{bmatrix}$$

$X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y}$

$\uparrow$ infinitely many solution.

Exercise: Consider the model

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, a$$

for some $a \geq 2$. Determine whether the following quantities are estimable:

1. μ
2. $\mu + \sum_{i=1}^{a} \alpha_i$
3. $(\mu + \alpha_2) - (\mu + \alpha_1)$

check if $\mu + \alpha_i$ is estimable, $i = 1, \ldots, a$.

① $\mu = \underset{\sim}{c}^T \underset{\sim}{b} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^T \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \end{pmatrix}$ is estimable if $\underset{\sim}{c} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \text{Col } X^T.$

$$\text{Col } X^T = \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} = \text{Col} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 1 \end{bmatrix}$$

$$X^T = \begin{bmatrix} \underset{\sim}{1}_{n_1}^T & \underset{\sim}{1}_{n_2}^T & \cdots & \underset{\sim}{1}_{n_a}^T \\ \underset{\sim}{1}_{n_1}^T & \underset{\sim}{0}^T & \cdots & \underset{\sim}{0}^T \\ \underset{\sim}{0}^T & \underset{\sim}{1}_{n_2}^T & \cdots & \underset{\sim}{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{\sim}{0}^T & \underset{\sim}{0}^T & & \underset{\sim}{1}_{n_a}^T \end{bmatrix}$$

Since $\underset{\sim}{c} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \notin \text{Col } X^T$, μ is not estimable.

② $\mu + \sum_{i=1}^{a} \alpha_i = \underset{\sim}{c}^T \underset{\sim}{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}^T \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \end{pmatrix}.$

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \overset{?}{\in} \text{Col} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 1 \end{bmatrix}$$

• set $a = 3$. Then $\mu + \alpha_1 + \alpha_2 + \alpha_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}^T \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$

$$X = \begin{bmatrix} \underset{\sim}{1}_{n_1} & \underset{\sim}{1}_{n_1} & & \\ \underset{\sim}{1}_{n_2} & & \underset{\sim}{1}_{n_2} & \\ \underset{\sim}{1}_{n_3} & & & \underset{\sim}{1}_{n_3} \end{bmatrix}$$

$$\text{Col } X^T = \text{Col} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Is $\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \in \text{Col} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Try solving this (see if it has a solution).

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}\ \underset{\sim}{x} = \underset{\overset{?}{\in} \underline{\text{Col } X^T}}{\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}}$$

Augment

$$\left[\begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right] \sim \left[\begin{array}{ccc:c} 0 & 0 & 0 & -2 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

$\Rightarrow$ No solution.

So we conclude $\mu + \sum_{i=1}^{a} \alpha_i$ is <u>not</u> estimable!

③ $$(\mu + \alpha_2) - (\mu + \alpha_1) = \underset{\underset{\sim}{c}}{\begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^{T}} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \end{pmatrix}$$

Do we have $\underset{\sim}{c} \in \text{Col } X^T$? Yes!

$$X^T = \begin{bmatrix} \underset{\sim}{1}_{n_1}^T & \underset{\sim}{1}_{n_2}^T & \cdots & \underset{\sim}{1}_{n_a}^T \\ \underset{\sim}{1}_{n_1}^T & \underset{\sim}{0}^T & \cdots & \underset{\sim}{0}^T \\ \underset{\sim}{0}^T & \underset{\sim}{1}_{n_2}^T & \cdots & \underset{\sim}{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{\sim}{0}^T & \underset{\sim}{0}^T & & \underset{\sim}{1}_{n_a}^T \end{bmatrix}$$

So $(\mu + \alpha_2) - (\mu + \alpha_1)$ <u>is</u> estimable.

$\underset{\sim}{c}^T \underset{\sim}{b}$

Least-squares estimator of an estimable contrast

The LS estimator of a contrast is $\mathbf{c}^T\hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is any vec. s.t. $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$.

Result (Properties of the LS estimator of an estimable contrast)

Let $\mathbf{c}^T\mathbf{b}$ be an estimable contrast. Then the LS estimator $\mathbf{c}^T\hat{\mathbf{b}}$

1. *is invariant to the choice of $\hat{\mathbf{b}}$ which satisfies $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$.*
2. *has expected value equal to $\mathbf{c}^T\mathbf{b}$ for all $\mathbf{b}$.*

See Res 3.2 and 3.3 of Monahan (2008).

Prove the results.

(1) Let $\underset{\sim}{c}^T \underset{\sim}{b}$ be estimable. Means $\underset{\sim}{c} \in \mathcal{C}(X^T)$.

Means $\exists$ $\underset{\sim}{a}$ such that $\underset{\sim}{c} = X^T \underset{\sim}{a}$.

Then let $\hat{\underset{\sim}{b}}_1$ satisfy $X^T X \hat{\underset{\sim}{b}}_1 = X^T \underset{\sim}{y}$

$\hat{\underset{\sim}{b}}_2$ satisfy $X^T X \hat{\underset{\sim}{b}}_2 = X^T \underset{\sim}{y}$.

Then

$$X^T X \hat{\underset{\sim}{b}}_1 = X^T X \hat{\underset{\sim}{b}}_2$$

$$\iff \quad X^T X \left(\hat{\underset{\sim}{b}}_1 - \hat{\underset{\sim}{b}}_2 \right) = \underset{\sim}{0}$$

$$\Rightarrow \quad \hat{\underset{\sim}{b}}_1 - \hat{\underset{\sim}{b}}_2 \in \text{Null } X^T X = \text{Null } X$$

$$\Rightarrow \quad X\left(\hat{\underset{\sim}{b}}_1 - \hat{\underset{\sim}{b}}_2 \right) = \underset{\sim}{0}.$$

$$\Rightarrow \quad X \hat{\underset{\sim}{b}}_1 = X \hat{\underset{\sim}{b}}_2 .$$

Now

$$\underset{\sim}{c}^T \hat{\underset{\sim}{b}}_1 = \left(X^T \underset{\sim}{a} \right)^T \hat{\underset{\sim}{b}}_1 = \underset{\sim}{a}^T X \hat{\underset{\sim}{b}}_1 = \underbrace{\underset{\sim}{a}^T X}_{\underset{\sim}{c}^T} \hat{\underset{\sim}{b}}_2 = \underset{\sim}{c}^T \hat{\underset{\sim}{b}}_2 .$$

② Let $\hat{\underset{\sim}{b}}$ satisfy $X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y}$.

$$\boxed{\left[X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y} \iff X \hat{\underset{\sim}{b}} = P_X \underset{\sim}{y} \right]}$$

$$X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y}$$

$$= X^T \left(P_X \underset{\sim}{y} + (I - P_X) \underset{\sim}{y} \right)$$

$$= X^T P_X \underset{\sim}{y} + \underbrace{X^T \underbrace{(I - P_X) \underset{\sim}{y}}_{\in (\text{Col } X)^\perp = \text{Null } X^T}}_{0}$$

$$= X^T P_X \underset{\sim}{y}$$

$$= X^T X (X^T X)^- X^T \underset{\sim}{y}$$

Cool result

$$\Rightarrow \quad X \hat{\underset{\sim}{b}} = X (X^T X)^- X^T \underset{\sim}{y} = P_X \underset{\sim}{y}$$

For estimable $\underset{\sim}{a}^T\underset{\sim}{b}$, we have

$$\mathbb{E}\,\underset{\sim}{c}^T\hat{\underset{\sim}{b}} = \mathbb{E}\left[(X^T\underset{\sim}{a})^T\hat{\underset{\sim}{b}}\right] = \mathbb{E}\left[\underset{\sim}{a}^T X\hat{\underset{\sim}{b}}\right] = \mathbb{E}\left[\underset{\sim}{a}^T P_X \underset{\sim}{Y}\right]$$

$\underset{\sim}{c} \in \text{Col}\, X^T$

$$= \underset{\sim}{a}^T P_X\, \mathbb{E}\,\underset{\sim}{Y}$$

$$= \underset{\sim}{a}^T P_X X\underset{\sim}{b}$$

$$= \underset{\sim}{a}^T X \underbrace{\boxed{(X^TX)^- X^T}}_{\text{gen inv. of } X} X\underset{\sim}{b}$$

$$= \underset{\sim}{a}^T X\underset{\sim}{b}$$

$$= \underset{\sim}{c}^T\underset{\sim}{b}$$

$\mathbb{E}\underset{\sim}{e} = \underset{\sim}{0}$

We can construct different models which yield the same predictions but in which the parameters have different interpretations.

$\mathbb{E}\underset{\sim}{y} = X\underset{\sim}{b}$ $\mathbb{E}\underset{\sim}{y} = W\underset{\sim}{d}$

Model reparameterization

Two linear models $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ and $\mathbf{y} = \mathbf{W}\mathbf{d} + \mathbf{e}$ are called *reparameterizations* of each other if $\operatorname{Col}\mathbf{X} = \operatorname{Col}\mathbf{W}$.

Exercise: Consider the three linear models

1. $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $i = 1, \dots, a$, $j = 1, \dots, n_i$.
2. $Y_{ij} = \alpha_i + \varepsilon_{ij}$, $i = 1, \dots, a$, $j = 1, \dots, n_i$.
3. $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $i = 1, \dots, a$, $j = 1, \dots, n_i$, where $\alpha_a = 0$.

Check whether these models are reparameterizations of each other.

① $$X_1 = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{1}_{n_2} & \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_a} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_a} \end{bmatrix} \qquad \mathbf{b}_1 = \begin{bmatrix} \mu \\ d_1 \\ \vdots \\ d_a \end{bmatrix}$$

② $$X_2 = \begin{bmatrix} \mathbf{1}_{n_1} & & & \\ & \mathbf{1}_{n_2} & & \\ & & \ddots & \\ & & & \mathbf{1}_{n_a} \end{bmatrix} \qquad \mathbf{b}_2 = \begin{bmatrix} d_1 \\ \vdots \\ d_a \end{bmatrix}$$

③ $$X_3 = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \cancel{\mathbf{0}} \\ \mathbf{1}_{n_2} & \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \cancel{\mathbf{0}} \\ \vdots & \vdots & \vdots & \ddots & \cancel{\vdots} \\ \mathbf{1}_{n_{a-1}} & \mathbf{0} & \mathbf{0} & \mathbf{1}_{n_{a-1}} & \cancel{\mathbf{0}} \\ \mathbf{1}_{n_a} & \mathbf{0} & \mathbf{0} & \cdots \mathbf{0} & \cancel{\mathbf{1}_{n_a}} \end{bmatrix} \qquad \mathbf{b}_3 = \begin{bmatrix} \mu \\ d_1 \\ \vdots \\ d_{a-1} \\ \cancel{D} \end{bmatrix}$$

$$X_3 = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0} & \cdots \\ \mathbf{1}_{n_2} & \mathbf{0} & \mathbf{1}_{n_2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \mathbf{1}_{n_{a-1}} & \mathbf{0} & \mathbf{0} & \mathbf{1}_{n_{a-1}} \\ \mathbf{1}_{n_a} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \qquad \mathbf{b}_3 = \begin{bmatrix} \mu \\ d_1 \\ \vdots \\ d_{a-1} \end{bmatrix}$$

$\underset{\sim}{y} = X\underset{\sim}{b} + \underset{\sim}{e}$

$X\hat{\underset{\sim}{b}} = P_X \underset{\sim}{y}$, ← "fitted values"

where $\hat{\underset{\sim}{b}}$ satisfies $X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y}$

$\underset{\sim}{y} = W\underset{\sim}{d} + \underset{\sim}{e}$

$W\hat{\underset{\sim}{d}} = P_W \underset{\sim}{y}$, ← "fitted values"

where $\hat{\underset{\sim}{d}}$ satisfies $W^T W \hat{\underset{\sim}{d}} = W^T \underset{\sim}{y}$.

Theorem (Difference of projection matrices)

Let $\mathbf{W}$ *and* $\mathbf{X}$ *be two matrices.*

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

$P_W = W(W^TW)^- W^T$

1. *If* $\mathrm{Col}\,\mathbf{X} = \mathrm{Col}\,\mathbf{W}$ *then* $\mathbf{P}_\mathbf{X} = \mathbf{P}_\mathbf{W}$.
2. *If* $\mathrm{Col}\,\mathbf{W} \subset \mathrm{Col}\,\mathbf{X}$ *then* $\mathbf{P}_\mathbf{X} - \mathbf{P}_\mathbf{W}$ *is the projection onto* $\mathrm{Col}((\mathbf{I} - \mathbf{P}_\mathbf{W})\mathbf{X})$.

See Thm 2.2 and 2.8 of Monahan (2008).

Prove the results.

① Let $\underline{\mathrm{Col}\, X = \mathrm{Col}\, W}$. Then

$$(P_X - P_W)^T (P_X - P_W) = P_X^T P_X - P_X^T P_W - P_W^T P_X + P_W^T P_W$$

$$= P_X - \underbrace{P_X P_W}_{P_W} - \underbrace{P_W P_X}_{P_X \text{ (likewise}} + P_W$$

$$= P_X - P_W - P_X + P_W$$

$$= 0$$

We have $\quad P_X \underbrace{P_W \underset{\sim}{z}}_{\substack{\in \text{Col}\, W, \\ \text{so } \in \text{Col}\, X}} = P_W \underset{\sim}{z} \quad \forall \underset{\sim}{z} \quad \Rightarrow \quad P_X P_W = P_W$

$A\underset{\sim}{x} = B\underset{\sim}{x} \quad \forall \underset{\sim}{x}$

$$\Rightarrow \quad P_X = P_W .$$

Result (Estimability in reparameterized models)

Consider the models $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ *and* $\mathbf{y} = \mathbf{W}\mathbf{d} + \mathbf{e}$*, where* $\mathbb{E}\mathbf{e} = \mathbf{0}$*. Suppose*

$$\operatorname{Col}\mathbf{W} = \operatorname{Col}\mathbf{X} \quad \text{with} \quad \mathbf{W} = \mathbf{X}\mathbf{T} \text{ and } \mathbf{X} = \mathbf{W}\mathbf{S}$$

and let $\hat{\mathbf{b}}$ *and* $\hat{\mathbf{d}}$ *satisfy* $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$ *and* $\mathbf{W}^T\mathbf{W}\hat{\mathbf{d}} = \mathbf{W}^T\mathbf{y}$.

$\underset{\sim}{c}^T \underset{\sim}{d} = \underset{\sim}{c}^T(S\underset{\sim}{b})$

Suppose $\mathbf{c}^T\mathbf{b}$ *is estimable in the model* $= \mathbf{X}\mathbf{b} + \mathbf{e}$*. Then* $\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(\mathbf{T}\mathbf{d})$ *and*

1. $\mathbf{c}^T(\mathbf{T}\mathbf{d})$ *is estimable in the model* $\mathbf{y} = \mathbf{W}\mathbf{d} + \mathbf{e}$.
2. $\mathbf{c}^T\mathbf{b}$ *has least squares estimator* $\mathbf{c}^T(\mathbf{T}\hat{\mathbf{d}})$.

See Res 3.4 and 3.5 of Monahan (2008).

~~**Prove the result.**~~

See the book.

$$Y_{ij} = \begin{cases} \mu + d_1 + \varepsilon_{1j}, & i=1 \\ \mu + d_2 + \varepsilon_{2j}, & i=2 \\ \mu + d_3 + \varepsilon_{3j}, & i=3 \end{cases}$$

$$Y_{ij} = \begin{cases} \tau_1 + \tau_2 + \varepsilon_{1j} & \text{if } i=1 \\ \tau_1 + \tau_3 + \varepsilon_{2j} & \text{if } i=2 \\ \tau_1 + \varepsilon_{3j} & \text{if } i=3 \end{cases}$$

Exercise: Consider the models $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ and $\mathbf{y} = \mathbf{W}\mathbf{d} + \mathbf{e}$, where

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

$$\mathbf{X}\mathbf{b} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} & \mathbf{0}_{n_3} & \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \quad \text{and} \quad \mathbf{W}\mathbf{d} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} \\ \mathbf{1}_{n_3} & \mathbf{0}_{n_3} & \mathbf{0}_{n_3} \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}$$

and $\mathbb{E}\mathbf{e} = \mathbf{0}$. Index the entries of $\mathbf{y}$ as Y_{ij}, $j = 1, \dots, n_i$, $i = 1, 2, 3$.

1. Show that $\mu + \alpha_3$ is estimable.
2. Give the matrix $\mathbf{T}$ such that $\mathbf{W} = \mathbf{X}\mathbf{T}$.
3. Give $\mu + \alpha_3$ in terms of τ_1, τ_2, and τ_3.
4. Show that τ_1, τ_2, and τ_3 are estimable.
5. Give the least-squares estimator of τ_1, τ_2, and τ_3 in terms of the entries of $\mathbf{y}$.
6. Give the least-squares estimator of $\mu + \alpha_3$ in terms of the entries of $\mathbf{y}$.
7. Give the matrix $\mathbf{S}$ such that $\mathbf{X} = \mathbf{W}\mathbf{S}$.
8. Give the parameters τ_1, τ_2, and τ_3 in terms of μ, α_1, α_2, and α_3.

Find $\hat{\underset{\sim}{d}}$ such that $W^T W \hat{\underset{\sim}{d}} = W^T \underset{\sim}{y}$.

Take $\hat{\tau}_1 = \underset{\sim}{c}_1^T \hat{\underset{\sim}{d}} \dots$

$$\mathbf{Xb} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} & \mathbf{0}_{n_3} & \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{bmatrix} \underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}}_{\underset{\sim}{b}}$$

① $\mu + \alpha_3$ estimable?
$$\underset{\sim}{c}^T \underset{\sim}{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}^T \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}.$$

Yes, because $\underset{\sim}{c} \in \text{Col } X^T = \text{Row } X$

②

$$W = X T$$

$$\underbrace{\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}}_{3 \times 3} = \underbrace{\underbrace{\boxed{\begin{matrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{matrix}}}_{3 \times 4} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{4 \times 3}}_{3 \times 3}$$

$$A\underset{\sim}{b} = [\underset{\sim}{a}_1 \cdots \underset{\sim}{a}_n] \underset{\sim}{b} = \sum_{i=1}^{n} \underset{\sim}{a}_i b_i$$

$$\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(\mathbf{Td})$$

③ $\mu + \alpha_3$ in terms of τ_1, τ_2, τ_3.

$$\mu + \alpha_3 = \underset{\sim}{c}^T \underset{\sim}{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}^T \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}^T \left(\underbrace{\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{4 \times 3} \underbrace{\begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}}_{3 \times 1}}_{4 \times 1} \right)$$

$$= \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}^T \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ 0 \end{pmatrix}$$

$$= \tau_1.$$

(4)

$$\mathbf{W}\mathbf{d} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} \\ \mathbf{1}_{n_3} & \mathbf{0}_{n_3} & \mathbf{0}_{n_3} \end{bmatrix} \overset{\underset{\sim}{d}}{\begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}}$$

τ_1, τ_2, τ_3 all estimable?

$\tau_1 = \underset{\sim}{c}_1^T \underset{\sim}{d} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}^T \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}$. Yes, because $\underset{\sim}{c}_1 \in \text{Row } W = \text{Col } W^T$

$\tau_2 = \underset{\sim}{c}_2^T \underset{\sim}{d} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}^T \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}$. Yes, $\underset{\sim}{c}_2 \in \text{Row } W$

$\tau_3 = \underset{\sim}{c}_3^T \underset{\sim}{d} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}^T \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}$. Yes, $\underset{\sim}{c}_3 \in \text{Row } W$.

(5) LS est. of τ_1, τ_2, τ_3.

Find solution to $W^T W \underset{\sim}{d} = W^T \underset{\sim}{y}$.

$$W^T W = \begin{bmatrix} n_1+n_2+n_3 & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{bmatrix} \qquad \underset{3\times n}{W^T} \underset{n\times 1}{\underset{\sim}{y}} = \underset{3\times 1}{\begin{bmatrix} n_1\bar{y}_{1\cdot} + n_2\bar{y}_{2\cdot} + n_3\bar{y}_{3\cdot} \\ n_1\bar{y}_{1\cdot} \\ n_2\bar{y}_{2\cdot} \end{bmatrix}}$$

$$\mathbf{W}\mathbf{d} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} \\ \mathbf{1}_{n_3} & \mathbf{0}_{n_3} & \mathbf{0}_{n_3} \end{bmatrix} \overset{\sim}{\begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}}$$

rank W = rank W^T

Col W^T = Col $W^T W$

$$
\left[\begin{array}{ccc|c}
n_1+n_2+n_3 & n_1 & n_2 & n_1\bar{y}_{1\cdot} + n_2\bar{y}_{2\cdot} + n_3\bar{y}_{3\cdot} \\
n_1 & n_1 & 0 & n_1\bar{y}_{1\cdot} \\
n_2 & 0 & n_2 & n_2\bar{y}_{2\cdot}
\end{array}\right]
$$

$$
\sim \left[\begin{array}{ccc|c}
n_3 & 0 & 0 & n_3\bar{y}_{3\cdot} \\
n_1 & n_1 & 0 & n_1\bar{y}_{1\cdot} \\
n_2 & 0 & n_2 & n_2\bar{y}_{2\cdot}
\end{array}\right]
$$

$$
\sim \left[\begin{array}{ccc|c}
1 & 0 & 0 & \bar{y}_{3\cdot} \\
1 & 1 & 0 & \bar{y}_{1\cdot} \\
1 & 0 & 1 & \bar{y}_{2}
\end{array}\right]
$$

$$
\sim \left[\begin{array}{ccc|c}
1 & 0 & 0 & \bar{y}_{3\cdot} \\
0 & 1 & 0 & \bar{y}_{1\cdot} - \bar{y}_{3\cdot} \\
0 & 0 & 1 & \bar{y}_{2\cdot} - \bar{y}_{3\cdot}
\end{array}\right]
$$

$$
\hat{\underset{\sim}{d}} = \begin{bmatrix} \bar{y}_{3\cdot} \\ \bar{y}_{1\cdot} - \bar{y}_{3\cdot} \\ \bar{y}_{2\cdot} - \bar{y}_{3\cdot} \end{bmatrix}
$$

$\hat{\tau}_1 = \bar{y}_{3\cdot}$, $\hat{\tau}_2 = \bar{y}_{1\cdot} - \bar{y}_{3\cdot}$ $\hat{\tau}_3 = \bar{y}_{2\cdot} - \bar{y}_{3\cdot}$

(c) $\mu + \alpha_3$ has LS estimator $\hat{\tau}_1 = \bar{y}_{3\cdot}$

⑦ Find S such that $X = WS$

$$\underset{X}{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}} = \underset{W}{\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}} \underset{\substack{3 \times 4 \\ \text{dim}}}{\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}}$$

⑧ Find τ_1, τ_2, τ_3 in terms of $\mu, \alpha_1, \alpha_2, \alpha_3$.

$$\tau_1 = \underset{\sim}{c}_1^T \underset{\sim}{d} = \underset{\sim}{c}_1^T (S \underset{\sim}{b}) = \underset{\sim}{c}_1^T \left[\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \right]$$

$$= \underset{\sim}{c}_1^T \begin{bmatrix} \mu + \alpha_3 \\ \alpha_1 - \alpha_3 \\ \alpha_2 - \alpha_3 \end{bmatrix}$$

$$=$$

$$\tau_1 = \mu + \alpha_3$$

$$\tau_2 = \underset{\sim}{c}_2^T (S \underset{\sim}{b}) = \alpha_1 - \alpha_3$$

$$\tau_3 = \underset{\sim}{c}_3^T (S \underset{\sim}{b}) = \alpha_2 - \alpha_3.$$

$X\underset{\sim}{b} = \underset{\sim}{y}$

$\|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}\| \leq \|\underset{\sim}{y} - X\underset{\sim}{b}\| \quad \forall \quad \underset{\sim}{b} \in \mathbb{R}^p$

$\iff$

$X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y}$

$\uparrow$

$\hat{\underset{\sim}{b}}$ is unique only when X has full column rank

1 Estimability

2 Reparameterization

3 Imposing conditions for a unique solution

4 Estimation in a restricted model

Recall this theorem from the previous lecture:

Theorem (Characterization of solutions to the normal equations)

The vector $\hat{\mathbf{b}}$ *is a solution to* $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ *iff there exists a vector* $\mathbf{z}$ *such that*

$$\hat{\mathbf{b}} = \boxed{(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}} + \boxed{(\mathbf{I} - (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X})\mathbf{z}.}$$

If $\mathbf{X}$ *has full-column rank, then* $\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ *is the unique solution.*

If $\mathbf{X}$ does not have full-column rank, the normal eqs do not have a unique solution.

If we *want* a unique solution, we can impose constraints of the form $\mathbf{C}\mathbf{b} = \mathbf{0}$. solve

$$\boxed{\begin{aligned} X^TX\,b &= X^Ty \\ Cb &= 0 \end{aligned}}$$

Then we augment the normal equations as $\begin{bmatrix} \mathbf{X}^T\mathbf{X} \\ \mathbf{C} \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{0} \end{bmatrix}$.

How can we choose $\mathbf{C}\mathbf{b}$ to make the augmented eqs have a unique solution?

Strategy: Take contrasts which are non-estimable and fix their values.

Recall: $\underset{\sim}{c}^T \underset{\sim}{b}$ is estimable $\iff$ $\underset{\sim}{c} \in \text{Col}\, X^T$

$X^T X \underset{\sim}{b} = X^T \underset{\sim}{y}$ has many solutions

Result (Choosing constraints for a unique solution)

Let $\mathbf{X}$ have rank $r < p$ and let $\mathbf{C}$ be a $(p - r) \times p$ matrix of which

1. *the rows are linearly independent and*
2. *each row defines a non-estimable contrast.*

I want rows of C which are not in $\text{Col}\, X^T = \text{Row}\, X$.

Then the augmented normal equations $\begin{bmatrix} \mathbf{X}^T\mathbf{X} \\ \mathbf{C} \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{0} \end{bmatrix}$ *have a unique solution.*

Rows of C cannot be linear combinations of the rows of X.

Prove the result.

$$\begin{bmatrix} X^T X \\ C \end{bmatrix} \underset{\sim}{b} = \begin{bmatrix} X^T y \\ \underset{\sim}{0} \end{bmatrix} \iff X^T X \underset{\sim}{b} = X^T \underset{\sim}{y} \text{ and } C\underset{\sim}{b} = \underset{\sim}{0}$$

$$\iff X\underset{\sim}{b} = P_X \underset{\sim}{y} \text{ and } C\underset{\sim}{b} = \underset{\sim}{0}$$

$$\iff \begin{bmatrix} X \\ C \end{bmatrix} \underset{\sim}{b} = P_X \underset{\sim}{y}.$$

Want to show that $\begin{bmatrix} X \\ C \end{bmatrix} \underset{\sim}{b} = P_X \underset{\sim}{y}$ has a unique solution.

I need $\begin{bmatrix} X \\ C \end{bmatrix}$ to have full-column rank.

$$\operatorname{rank} \begin{bmatrix} X \\ C \end{bmatrix} = \dim \operatorname{Col} \begin{bmatrix} X \\ C \end{bmatrix} = \dim \operatorname{Row} \begin{bmatrix} X \\ C \end{bmatrix}$$

$\dim \operatorname{Row} X = r$

$\dim \operatorname{Row} C = p-r$

Because rows of C are not in $\operatorname{Col} X^T$,

$$\dim \operatorname{Row} \begin{bmatrix} X \\ C \end{bmatrix} = r + p - r = p.$$

So $\begin{bmatrix} X \\ C \end{bmatrix}$ has full column rank.

$Y_i = \mu + \alpha_i + \varepsilon_{ij} \qquad i = 1, 2, 3, \qquad j = 1, \dots, n_i$

Unique solution to $X^T X \underset{\sim}{b} = X^T \underset{\sim}{Y}$?

Exercise: In the model $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$, where

$p = 4$

$r = 3$

$$\mathbf{Xb} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} & \mathbf{0}_{n_3} & \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix},$$

the Normal equations do not have a unique solution.

1. Give a matrix $\mathbf{C}$ such that the augmented normal eqs have a unique solution. ✓
2. Give the unique solution in terms of the entries of $\mathbf{y}$.

(1) $\begin{bmatrix} X^T X \\ C \end{bmatrix} \underset{\sim}{b} = \begin{bmatrix} X^T \underset{\sim}{Y} \\ \underset{\sim}{0} \end{bmatrix}$

$\boxed{C \underset{\sim}{b} = \underset{\sim}{0}}$ ← constraint

C has dimension 1×4

Try $C = [1 \ 0 \ 0 \ 0]$

Then $C\underset{\sim}{b} = [1 \ 0 \ 0 \ 0] \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \mu$, is not estimable.

Impose $C\underset{\sim}{b} = \underset{\sim}{0}$, which means $[1\ 0\ 0\ 0]\begin{pmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix} = \mu = 0.$

② $$\begin{bmatrix} X^TX \\ 1\ 0\ 0\ 0 \end{bmatrix}\begin{pmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix} = \begin{bmatrix} X^T\underset{\sim}{y} \\ 0 \end{bmatrix},$$

$$X^TX = \begin{bmatrix} n_1+n_2+n_3 & n_1 & n_2 & n_3 \\ n_1 & n_1 & 0 & 0 \\ n_2 & 0 & n_2 & 0 \\ n_3 & 0 & 0 & n_3 \end{bmatrix} \qquad X^T\underset{\sim}{y} = \begin{bmatrix} Y_{\cdot\cdot} \\ Y_{1\cdot} \\ Y_{2\cdot} \\ Y_{3\cdot} \end{bmatrix}$$

$$X = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} & \mathbf{0}_{n_3} & \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{bmatrix}$$

Augmented Normal eqs are

$$\begin{bmatrix} n_1+n_2+n_3 & n_1 & n_2 & n_3 \\ n_1 & n_1 & 0 & 0 \\ n_2 & 0 & n_2 & 0 \\ n_3 & 0 & 0 & n_3 \\ 1 & 0 & 0 & 0 \end{bmatrix}\begin{pmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix} = \begin{bmatrix} Y_{\cdot\cdot} \\ Y_{1\cdot} \\ Y_{2\cdot} \\ Y_{3\cdot} \\ 0 \end{bmatrix}$$

$$\left[\begin{array}{cccc|c} n_1+n_2+n_3 & n_1 & n_2 & n_3 & Y_{\cdot\cdot} \\ n_1 & n_1 & 0 & 0 & Y_{1\cdot} \\ n_2 & 0 & n_2 & 0 & Y_{2\cdot} \\ n_3 & 0 & 0 & n_3 & Y_{3\cdot} \\ 1 & 0 & 0 & 0 & 0 \end{array}\right] \sim \left[\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 0 \\ n_1 & n_1 & 0 & 0 & Y_{1\cdot} \\ n_2 & 0 & n_2 & 0 & Y_{2\cdot} \\ n_3 & 0 & 0 & n_3 & Y_{3\cdot} \\ 1 & 0 & 0 & 0 & 0 \end{array}\right]$$

$\Rightarrow \quad \hat{\mu} = 0$

$\hat{\alpha}_1 = \bar{y}_{1\cdot}, \quad \hat{\alpha}_2 = \bar{y}_{2\cdot}, \quad \hat{\alpha}_3 = \bar{y}_{3\cdot}$

Result (Replace $\mathbf{C}$ with $\mathbf{C}^T\mathbf{C}$)

For $\mathbf{C}$ *constructed as in the previous theorem, we have*

$$\overset{(A)}{\begin{bmatrix} \mathbf{X}^T\mathbf{X} \\ \mathbf{C} \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{0} \end{bmatrix}} \iff \overset{(B)}{\begin{bmatrix} \mathbf{X}^T\mathbf{X} \\ \mathbf{C}^T\mathbf{C} \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{0} \end{bmatrix}} \iff \overset{(C)}{\boxed{(\mathbf{X}^T\mathbf{X} + \mathbf{C}^T\mathbf{C})\mathbf{b} = \mathbf{X}^T\mathbf{y}.}}$$

See Lem 3.1 of Monahan (2008).

Prove the result.

(A) $\Rightarrow$ (B) $\quad \begin{bmatrix} X^TX \\ C \end{bmatrix} \underset{\sim}{b} = \begin{bmatrix} X^T\underset{\sim}{y} \\ \underset{\sim}{0} \end{bmatrix} \Rightarrow X^TX\underset{\sim}{b} = X^T\underset{\sim}{y}$ and $C\underset{\sim}{b} = \underset{\sim}{0} \Rightarrow X^TX\underset{\sim}{b} = X^T\underset{\sim}{y}$, $C^TC\underset{\sim}{b} = 0 \Rightarrow \begin{bmatrix} X^TX \\ C^TC \end{bmatrix} \underset{\sim}{b} = \begin{bmatrix} X^T\underset{\sim}{y} \\ \underset{\sim}{0} \end{bmatrix}$

(B) $\Rightarrow$ (A) $\quad \begin{bmatrix} X^TX \\ C^TC \end{bmatrix} \underset{\sim}{b} = \begin{bmatrix} X^T\underset{\sim}{y} \\ \underset{\sim}{0} \end{bmatrix} \Rightarrow X^TX\underset{\sim}{b} = X^T\underset{\sim}{y}$, $C^TC\underset{\sim}{b} = \underset{\sim}{0}$. Can show $\text{Null}\, C^TC = \text{Null}\, C$. So $C^TC\underset{\sim}{b} = \underset{\sim}{0} \Rightarrow C\underset{\sim}{b} = \underset{\sim}{0}$.

$\Rightarrow X^TX\underset{\sim}{b} = X^T\underset{\sim}{y}$, $C\underset{\sim}{b} = \underset{\sim}{0} \Rightarrow \begin{bmatrix} X^TX \\ C \end{bmatrix} \underset{\sim}{b} = \begin{bmatrix} X^T\underset{\sim}{y} \\ \underset{\sim}{0} \end{bmatrix}$.

(B) $\Rightarrow$ (C)

$$\begin{bmatrix} X^TX \\ C^TC \end{bmatrix} \underset{\sim}{b} = \begin{bmatrix} X^T\underset{\sim}{y} \\ \underset{\sim}{0} \end{bmatrix} \Rightarrow \begin{matrix} X^TX\underset{\sim}{b} = X^T\underset{\sim}{y} \\ \text{and} \quad C^TC\underset{\sim}{b} = \underset{\sim}{0} \end{matrix} \Rightarrow X^TX\underset{\sim}{b} + C^TC\underset{\sim}{b} = X^T\underset{\sim}{y}$$

$$\Rightarrow (X^TX + C^TC)\underset{\sim}{b} = X^T\underset{\sim}{y}.$$

(C) $\Rightarrow$ (B)

$$(X^TX + C^TC)\underset{\sim}{b} = X^T y \Rightarrow C^TC\underset{\sim}{b} = X^T\underset{\sim}{y} - X^TX\underset{\sim}{b}$$

$$\Rightarrow \underbrace{C^TC\underset{\sim}{b}}_{\substack{\in \text{Col } C^T \\ \shortparallel \\ \text{Row } C}} = \underbrace{X^T(\underset{\sim}{y} - X\underset{\sim}{b})}_{\in \text{Col } X^T} = \underset{\sim}{0}.$$

$$\{\text{Row } C\} \cap \{\text{Col } X^T\} = \{\underset{\sim}{0}\}$$

$$\Rightarrow \quad C^TC\underset{\sim}{b} = \underset{\sim}{0}$$

$$\text{and} \quad X^T(\underset{\sim}{y} - X\underset{\sim}{b}) = \underset{\sim}{0} \iff X^TX\underset{\sim}{b} = X^T\underset{\sim}{y}$$

$$\Rightarrow \begin{bmatrix} X^TX \\ C^TC \end{bmatrix} \underset{\sim}{b} = \begin{bmatrix} X^T\underset{\sim}{y} \\ \underset{\sim}{0} \end{bmatrix}.$$

$$(\underset{\sim}{x}^T x + c^T c)\, \underset{\sim}{b} = x^T \underset{\sim}{y}$$

Theorem (Constraint-augmented Normal equations)

Let $\mathbf{X}$ *be an* $n \times p$ *with rank* $r < p$ *and let* $\mathbf{C}$ *be an* $(p-r) \times p$ *matrix with rank* $p - r$ *of which each row defines a non-estimable contrast. Then:*

1. $(\mathbf{X}^T\mathbf{X} + \mathbf{C}^T\mathbf{C})$ *is nonsingular.*
2. $(\mathbf{X}^T\mathbf{X} + \mathbf{C}^T\mathbf{C})^{-1}\mathbf{X}^T\mathbf{y}$ *is the unique solution to* $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ *and* $\mathbf{C}\mathbf{b} = \mathbf{0}$.
3. $(\mathbf{X}^T\mathbf{X} + \mathbf{C}^T\mathbf{C})^{-1}$ *is a generalized inverse of* $\mathbf{X}^T\mathbf{X}$.
4. $\mathbf{C}(\mathbf{X}^T\mathbf{X} + \mathbf{C}^T\mathbf{C})^{-1}\mathbf{X}^T = \mathbf{0}$.
5. $\mathbf{C}(\mathbf{X}^T\mathbf{X} + \mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T = \mathbf{I}$.

See Res 3.6 of Monahan (2008).

Means we get unique solution as

$$\hat{\underset{\sim}{b}} = (x^T x + c^T c)^{-1} x^T \underset{\sim}{y}.$$

Prove 1 and 2.

$$\underset{\sim}{c}^T \underset{\sim}{b} = \overbrace{\begin{bmatrix} 0 & n_1 & n_2 & n_3 \end{bmatrix}}^{C} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = n_1\alpha_1 + n_2\alpha_2 + n_3\alpha_3$$

Exercise: Let $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ for $j = 1, \ldots, n_i$, $i = 1, \ldots, a$. Find the least-squares estimator of each parameter under the constraint $\sum_{i=1}^{a} n_i\alpha_i = 0$.

$$\begin{bmatrix} X^T X \\ C \end{bmatrix} = \begin{bmatrix} n_1+n_2+n_3 & n_1 & n_2 & n_3 \\ n_1 & n_1 & 0 & 0 \\ n_2 & 0 & n_2 & 0 \\ n_3 & 0 & 0 & n_3 \\ \boxed{0 \quad n_1 \quad n_2 \quad n_3} \end{bmatrix} \qquad \begin{bmatrix} X^T \underset{\sim}{y} \\ 0 \end{bmatrix} = \begin{bmatrix} Y_{\cdot\cdot} \\ Y_{1\cdot} \\ Y_{2\cdot} \\ Y_{3\cdot} \\ 0 \end{bmatrix}$$

$$X = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} & \mathbf{0}_{n_3} & \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{bmatrix}$$

$$\begin{bmatrix} n_1+n_2+n_3 & n_1 & n_2 & n_3 & y_{\cdot\cdot} \\ n_1 & n_1 & 0 & 0 & y_{1\cdot} \\ n_2 & 0 & n_2 & 0 & y_{2\cdot} \\ n_3 & 0 & 0 & n_3 & y_{3\cdot} \\ 0 & n_1 & n_2 & n_3 & 0 \end{bmatrix} \sim \begin{bmatrix} n_1+n_2+n_3 & 0 & 0 & 0 & y_{\cdot\cdot} \\ n_1 & n_1 & 0 & 0 & y_{1\cdot} \\ n_2 & 0 & n_2 & 0 & y_{2\cdot} \\ n_3 & 0 & 0 & n_3 & y_{3\cdot} \\ 0 & n_1 & n_2 & n_3 & 0 \end{bmatrix}$$

$$\sim \begin{bmatrix} n_1+n_2+n_3 & 0 & 0 & 0 & y_{\cdot\cdot} \\ 0 & n_1 & 0 & 0 & y_{1\cdot} - \frac{n_1}{n_1+n_2+n_3} y_{\cdot\cdot} \\ 0 & 0 & n_2 & 0 & y_{2\cdot} - \frac{n_2}{n_1+n_2+n_3} y_{\cdot\cdot} \\ 0 & 0 & 0 & n_3 & y_{3\cdot} - \frac{n_3}{n_1+n_2+n_3} y_{\cdot\cdot} \\ 0 & n_1 & n_2 & n_3 & 0 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & 0 & 0 & \bar{y}_{\cdot\cdot} \\ 0 & 1 & 0 & 0 & \bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot} \\ 0 & 0 & 1 & 0 & \bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot} \\ 0 & 0 & 0 & 1 & \bar{y}_{3\cdot} - \bar{y}_{\cdot\cdot} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\hat{\mu} = \bar{y}_{\cdot\cdot}$, $\hat{\alpha}_1 = \bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot}$

$\hat{\alpha}_2 = \bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot}$

$\hat{\alpha}_3 = \bar{y}_{3\cdot} - \bar{y}_{\cdot\cdot}$

```
# set up parameters
a <- 4 # set number of 'treatment groups'
mu <- rnorm(1) # generate mu value
alpha <- rnorm(a) # generate alpha values
nn <- pmax(2,rpois(a,3)) # generate sample sizes no smaller than 2

# build design matrix
X <- matrix(0,nrow = sum(nn),ncol = a + 1)
trt <- numeric(sum(nn))
k <- 1
for(i in 1:a){
  ind <- k:(k - 1 + nn[i])
  X[ind,c(1,i+1)] <- 1
  trt[ind] <- i
  k <- k + nn[i]
}

# generate some Y values
b <- c(mu,alpha)
e <- rnorm(sum(nn))
Y <- as.numeric(X %*% b) + e

# compute constrained estimator
Cmat <- matrix(c(0,nn),nrow = 1)
b_hat <- solve(t(X) %*% X + t(Cmat) %*% Cmat) %*% t(X) %*% Y
b_hat

# check earlier work
mean(Y)
for(i in 1:a) print(mean(Y[trt == i]) - mean(Y))
```

$(X^T X + C^T C)^{-1} X^T \underset{\sim}{Y}$

We may have a reason to place restrictions on the parameters of $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$.

We consider placing restrictions on $\mathbf{b}$ of the form $\mathbf{P}^T\mathbf{b} = \delta$.

Can use this to form hypothesis tests or to impose structure on $\mathbf{b}$.

We will called the model $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ under $\mathbf{P}^T\mathbf{b} = \delta$ the *restricted model*.

Estimability in the restricted model

A contrast $\mathbf{c}^T\mathbf{b}$ is *estimable in the restricted model* if there exists a scalar a_0 and a vector $\mathbf{a}$ such that $\mathbb{E}[a_0 + \mathbf{a}^T\mathbf{y}] = \mathbf{c}^T\mathbf{b}$ for all $\mathbf{b}$ satisfying $\mathbf{P}^T\mathbf{b} = \delta$.

Result (What contrasts are estimable in the restricted model?)

A contrast $\mathbf{c}^T\mathbf{b}$ *is estimable in the restricted model if and only if* $\mathbf{c} \in \text{Col}\,[\mathbf{X}^T\ \mathbf{P}]$.

See Res 3.7 of Monahan (2008).

Prove the result. [see book]

Can find the best approx. $\mathbf{X}\hat{\mathbf{b}}$ to $\mathbf{y}$ subject to $\mathbf{P}^T\hat{\mathbf{b}} = \boldsymbol{\delta}$ with Lagrange multipliers.

Set $Q(\mathbf{b}, \mathbf{u}) = \frac{1}{2}\|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 + \mathbf{u}^T(\mathbf{P}^T\mathbf{b} - \boldsymbol{\delta})$ and set derivs wrt $\mathbf{b}$ and $\mathbf{u}$ to $\mathbf{0}$.

This gives the restricted normal equations (RNEs). Derive

Result (Consistency of restricted normal equations)

The restricted normal equations $\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \boldsymbol{\delta} \end{bmatrix}$ *are consistent.*

See Res 3.8 of Monahan (2008).

Prove the result. [See book]

$$\frac{\partial}{\partial \underset{\sim}{b}} Q(\underset{\sim}{b}, \underset{\sim}{u}) = -X^T(\underset{\sim}{y} - X\underset{\sim}{b}) + P\underset{\sim}{u} \overset{set}{=} \underset{\sim}{0}$$

$$\frac{\partial}{\partial \underset{\sim}{u}} Q(\underset{\sim}{b}, \underset{\sim}{u}) = P^T\underset{\sim}{b} - \underset{\sim}{\delta} \overset{set}{=} \underset{\sim}{0}$$

$$\Rightarrow \quad -X^T\underset{\sim}{y} + X^TX\underset{\sim}{b} + P\underset{\sim}{u} = \underset{\sim}{0}, \qquad P^T\underset{\sim}{b} = \underset{\sim}{\delta}$$

$$\Rightarrow \quad X^T X b + P m = X^T y \qquad \Rightarrow \quad \begin{bmatrix} X^T X & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} X^T y \\ \delta \end{bmatrix}$$

$$P^T b = \delta$$

Result (RNE solution gives best approximation in restricted model)

Let $\hat{\mathbf{b}}_H$ *and* $\hat{\mathbf{u}}$ *be any vectors such that*

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}}_H \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \delta \end{bmatrix}.$$

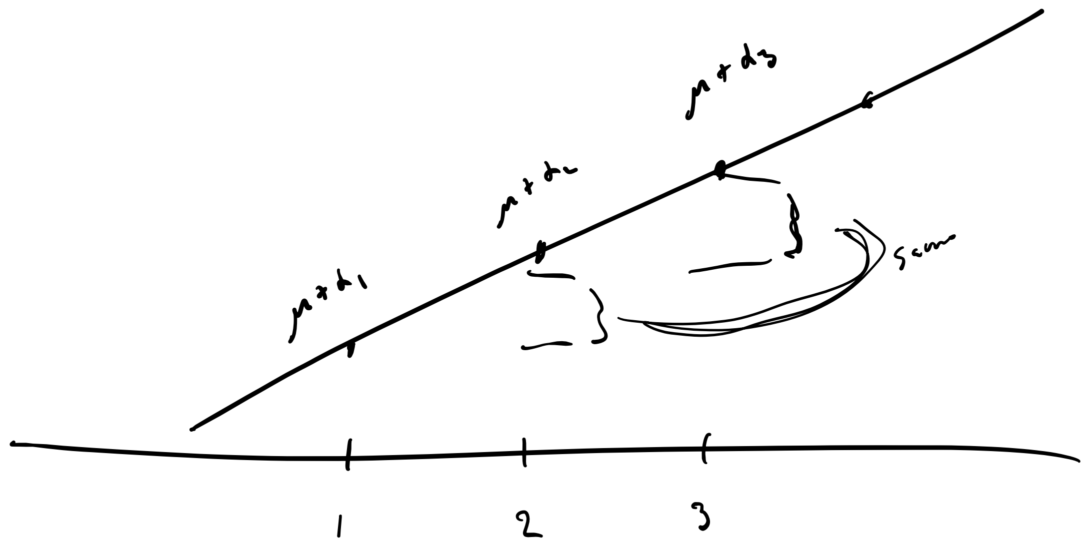
Then $\boxed{\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}_H\| \leq \|\mathbf{y} - \mathbf{X}\mathbf{b}\| \text{ for all } \mathbf{b} \text{ such that } \mathbf{P}^T\mathbf{b} = \delta.}$

See Res 3.9 of Monahan (2008).

Prove the result. Show.

$$\mu + d_3 - (\mu + d_2) = \mu + d_2 - (\mu + d_1)$$

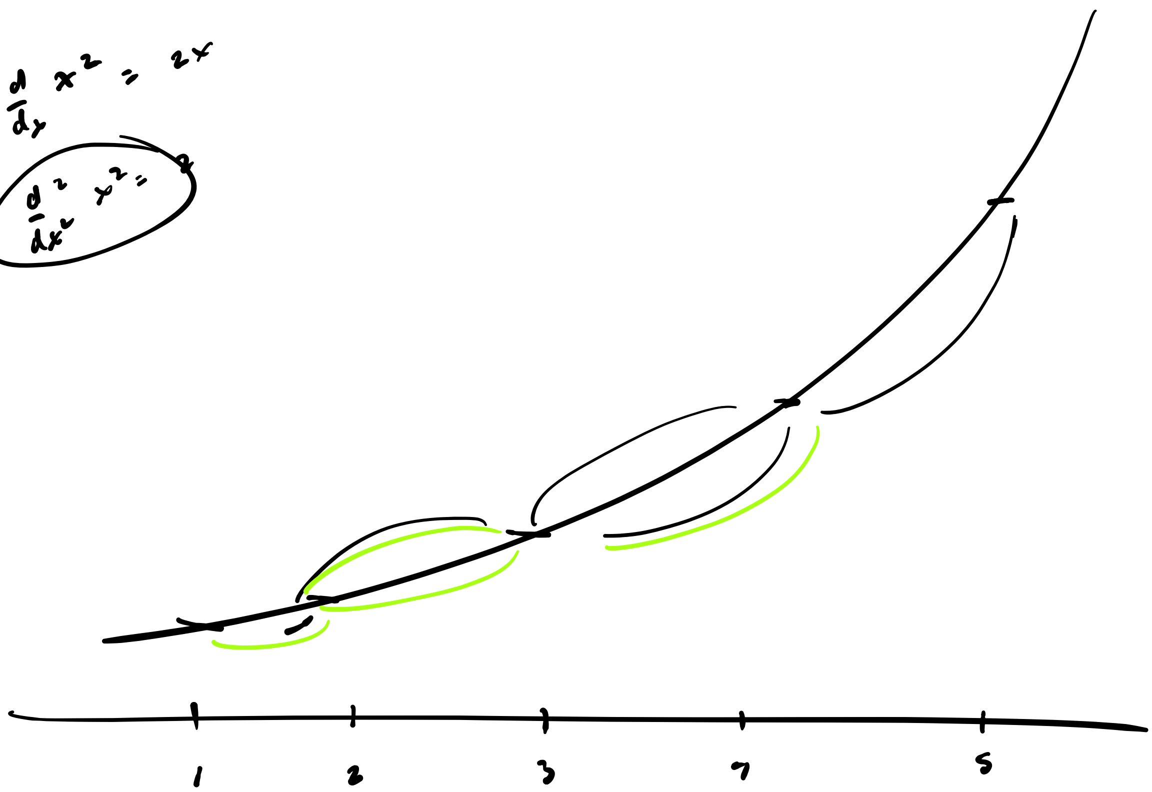
$$\boxed{d_3 - d_2 = d_2 - d_1}$$



$$\frac{d}{dx} x^2 = 2x$$

$$\frac{d^2}{dx^2} x^2 = 2$$

quadratic



$$\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{pmatrix}$$

Exercise: Let $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $j = 1, \dots, n_i$, $i = 1, 2, 3, 4, 5$.

Consider the constraints

$$\alpha_4 - 3\alpha_3 + 3\alpha_2 - \alpha_1 = 0$$

$$(\alpha_4 - \alpha_3) - (\alpha_3 - \alpha_2) = (\alpha_3 - \alpha_2) - (\alpha_2 - \alpha_1)$$

$$(\alpha_5 - \alpha_4) - (\alpha_4 - \alpha_3) = (\alpha_4 - \alpha_3) - (\alpha_3 - \alpha_2)$$

$$\alpha_5 - 3\alpha_4 + 3\alpha_3 - \alpha_2 = 0$$

1. Interpret these constraints.
2. Express the constraints as $\mathbf{P}^T\mathbf{b} = \boldsymbol{\delta}$.
3. Check whether μ is estimable in the restricted model. at home
4. Check whether $\mu + \alpha_i$, $i = 1, \dots, 5$ are estimable in the restricted model. yes
5. Is there a unique solution to the RNEs? ← No unique solution! at home.

(2) $$P^T = \begin{bmatrix} 0 & -1 & 3 & -3 & 1 & 0 \\ 0 & 0 & -1 & 3 & -3 & 1 \end{bmatrix}$$

$$\begin{bmatrix} X^T X & P \\ P^T & 0 \end{bmatrix}$$

$$\begin{bmatrix} X^T \underset{\sim}{y} \\ 0 \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ y_{3.} \\ 0 \\ 0 \end{bmatrix}$$

$$\left[\begin{array}{cccccc|cc}
n_1+n_2+n_3+n_4+n_5 & n_1 & n_2 & n_3 & n_4 & n_5 & & \\
n_1 & n_1 & 0 & 0 & 0 & 0 & & \\
n_2 & 0 & n_2 & 0 & 0 & 0 & \multicolumn{2}{c}{P} \\
n_3 & 0 & 0 & n_3 & 0 & 0 & & \\
n_4 & 0 & 0 & 0 & n_4 & 0 & & \\
n_5 & 0 & 0 & 0 & 0 & n_5 & & \\
0 & -1 & 3 & -3 & 1 & 0 & 0 & 0 \\
0 & 0 & -1 & 3 & -3 & 1 & 0 & 0
\end{array}\right]$$

```r
# set up parameters
a <- 8 # choose a number of 'treatment groups'
mu <- 2 # generate mu value
alpha <- (.25)*c(1:a - a/2)^2 + c(1:a) # generate quadratic effect
nn <- pmax(2,rpois(a,5)) # generate sample sizes no smaller than 2

# build design matrix and vector of treatment assignments
X <- matrix(0,nrow = sum(nn),ncol = a + 1)
trt <- numeric(sum(nn))
k <- 1
for(i in 1:a){
  ind <- k:(k - 1 + nn[i])
  X[ind,c(1,i+1)] <- 1
  trt[ind] <- i
  k <- k + nn[i]
}

# generate some Y values
b <- c(mu,alpha)
e <- rnorm(sum(nn))
Y <- as.numeric(X %*% b) + e

# construct restriction matrix
Pt <- matrix(0,nrow = a - 3,ncol = a+1)
for(j in 1:nrow(Pt)) Pt[j,(1+j):(1+j+3)] <- c(-1,3,-3,1)
P <- t(Pt)
```

$$P^T = \begin{bmatrix} 0 & -1 & 3 & -3 & 1 & 0 \\ 0 & 0 & -1 & 3 & -3 & 1 \end{bmatrix}$$

$$\begin{pmatrix} X^T X & P \\ P^T & 0 \end{pmatrix}$$

```
# set up RNEs
M <- rbind(cbind(t(X) %*% X,P),cbind(Pt,matrix(0,nrow(Pt),ncol(P))))
v <- c(t(X) %*% Y,rep(0,nrow(Pt)))
```

$$\begin{bmatrix} X^T y \\ \underset{\sim}{\delta} \end{bmatrix}$$

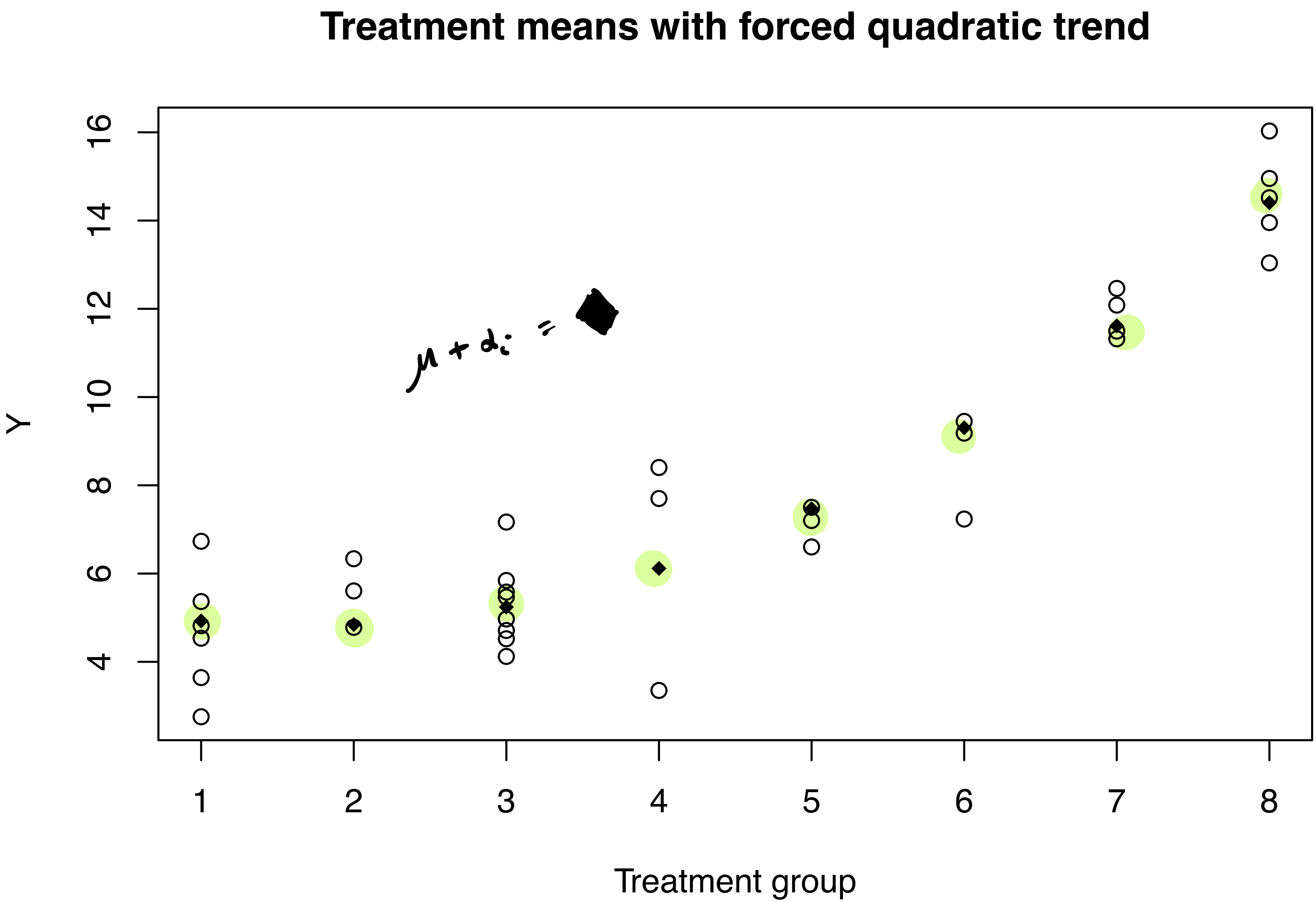
```
# obtain a generalized inverse for M
r <- a + a - 3
M_svd <- svd(M)
M_ginv <- M_svd$v[,1:r] %*% diag(1/M_svd$d[1:r]) %*% t(M_svd$u[,1:r])

# obtain a solution to the RNEs
bu_hat <- M_ginv %*% v
b_hat <- bu_hat[1:(a+1)]

# estimate contrasts of interest
Cmat <- cbind(rep(1,a),diag(a))
trt_means <- Cmat %*% b_hat
```

$\mu + a_i$ $\qquad$ $\underset{\sim}{c}^T \underset{\sim}{b}$

```
# make a plot
plot(trt_means,pch = 18,
     ylim = range(Y),
     ylab = "Y",
     xlab = "Treatment group",
     main = "Treatment means with forced quadratic trend")
points(Y~trt)
```


Treatment means with forced quadratic trend
$\mu + \alpha_i =$
Y
4
6
8
10
12
14
16
1
2
3
4
5
6
7
8
Treatment group

Monahan, J. F. (2008). *A primer on linear models*. CRC Press.