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.

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

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* $\mathbf{c} \in \operatorname{Col} \mathbf{X}^T$.

See Res 3.1 of Monahan (2008).

Prove the result.

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. $(\mu + \alpha_2) - (\mu + \alpha_1)$

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.

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

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.

Theorem (Difference of projection matrices)

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

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

See Thm 2.2 and 2.8 of Monahan (2008).

Prove the results.

Result (Estimability in reparameterized models)

Consider the models $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ *and* $\mathbf{y} = \mathbf{Wd} + \mathbf{e}$, *where* $\mathbb{E}\mathbf{e} = \mathbf{0}$. *Suppose*

$$\text{Col}\,\mathbf{W} = \text{Col}\,\mathbf{X} \quad \textit{with} \quad \mathbf{W} = \mathbf{XT} \textit{ and } \mathbf{X} = \mathbf{WS}$$

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

1. *If* $\mathbf{c}^T\mathbf{b}$ *is estimable in* $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ *then* $\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(\mathbf{Td})$, *and* $\mathbf{c}^T(\mathbf{Td})$ *is estimable in* $\mathbf{y} = \mathbf{Wd} + \mathbf{e}$ *with least-squares estimator* $\mathbf{c}^T(\mathbf{T}\hat{\mathbf{d}})$.
2. *If* $\mathbf{c}^T\mathbf{d}$ *is estimable in* $\mathbf{y} = \mathbf{Wd} + \mathbf{e}$ *then* $\mathbf{c}^T\mathbf{d} = \mathbf{c}^T(\mathbf{Sb})$, *and* $\mathbf{c}^T(\mathbf{Sb})$ *is estimable in* $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ *with least-squares estimator* $\mathbf{c}^T(\mathbf{S}\hat{\mathbf{b}})$.

See Res 3.4 and 3.5 of Monahan (2008).

Prove the result.

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

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

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.

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*

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

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 $\mathbf{C}\mathbf{b} = \mathbf{0}$.

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.

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.*

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.*

Prove the result.

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

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

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 $\mathbf{y}$.

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

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

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

See Lem 3.1 of Monahan (2008).

Prove the result.

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).

Prove 1 and 2.

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

```
# 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))
```

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 $\mathbf{P}^T\mathbf{b} = \boldsymbol{\delta}$.

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 $\mathbf{P}^T\mathbf{b} = \boldsymbol{\delta}$ 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 $\mathbf{P}^T\mathbf{b} = \boldsymbol{\delta}$.

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.

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

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

This gives the restricted normal equations (RNEs).

Result (Consistency of restricted normal equations)

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

See Res 3.8 of Monahan (2008).

Prove the result.

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

See Res 3.9 of Monahan (2008).

Prove the result.

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

Consider the constraints

$$\begin{aligned}(\alpha_4 - \alpha_3) - (\alpha_3 - \alpha_2) &= (\alpha_3 - \alpha_2) - (\alpha_2 - \alpha_1)\\ (\alpha_5 - \alpha_4) - (\alpha_4 - \alpha_3) &= (\alpha_4 - \alpha_3) - (\alpha_3 - \alpha_2)\end{aligned}$$

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.
4. Check whether $\mu + \alpha_i$, $i = 1, \dots, 5$ are estimable in the restricted model.
5. Is there a unique solution to the RNEs?

```
# 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)
```

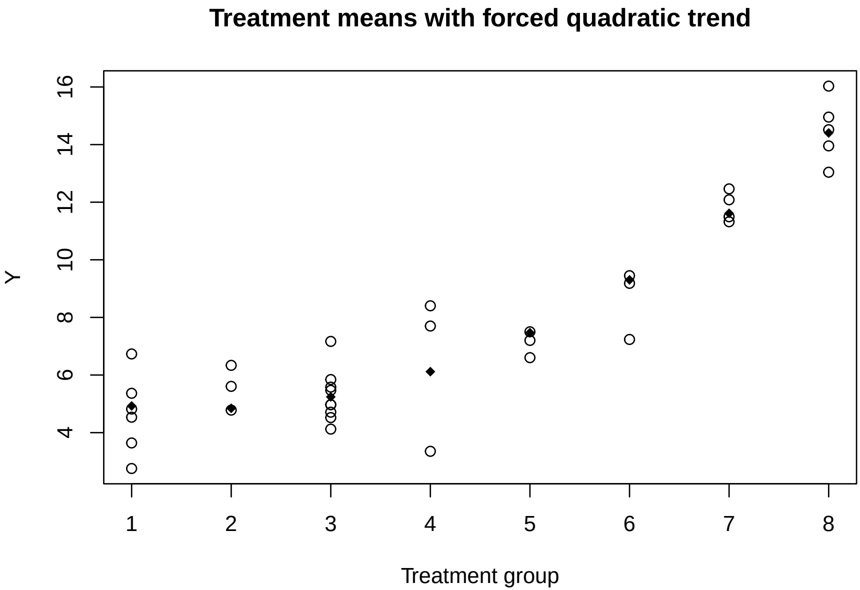
```
# set up RNEs
M <- rbind(cbind(t(X) %*% X,P),cbind(Pt,matrix(0,nrow(Pt),ncol(P))))
v <- c(t(X) %*% Y,rep(0,nrow(Pt)))

# 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

# 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
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.