

STAT 640: Homework 6

Due **Wednesday, March 9, 11:59pm MT** on the course Canvas webpage. Please follow the homework guidelines on the syllabus.

Name: Hannah Butler

Problem 1

Suppose that $\hat{\beta}_1$ and $\hat{\beta}_2$ are two *distinct* least squares estimators of β . Show that there are infinitely many combinations of $\hat{\beta}_1$ and $\hat{\beta}_2$ that are also LSEs of β .

Answer: If $\text{rank}(\mathbf{X}) = p$, then \mathbf{X} is full rank and there can not be 2 distinct least squares estimators of β (proposition 4.3). So $\text{rank}(\mathbf{X})$ must be less than p . This means that there exists $\mathbf{u} \neq \mathbf{0}$ in $\mathcal{N}(\mathbf{X})$. Then $a\mathbf{u}$ is in $\mathcal{N}(\mathbf{X})$ for any $a \in \mathbb{R}$. Since $\hat{\beta}_1$ is an LSE for β , $\mathbf{X}\hat{\beta}_1 = \mathbf{P}_\mathbf{X}\mathbf{Y}$ (proposition 4.1). So $\mathbf{X}(\hat{\beta}_1 + a\mathbf{u}) = \mathbf{X}\hat{\beta}_1 + a\mathbf{X}\mathbf{u} = \mathbf{X}\hat{\beta}_1 + \mathbf{0} = \mathbf{X}\hat{\beta}_1 = \mathbf{P}_\mathbf{X}\mathbf{Y}$. Therefore, $\hat{\beta}_1 + a\mathbf{u}$ must be a distinct LSE of β , for all $a \in \mathbb{R}$.

Problem 2

Consider the linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Assume the first column of \mathbf{X} is $\mathbf{x}_1 = \mathbf{1}$ and that $\text{rank}(\mathbf{X}) = p$ (i.e., \mathbf{X} is full-rank).

- a. Show that $\sum_{i=1}^n (Y_i - \hat{Y}_i) = 0$.

Answer:

$$\begin{aligned} \sum_{i=1}^n (Y_i - \hat{Y}_i) &= (\mathbf{Y} - \hat{\mathbf{Y}})^T \mathbf{1} \\ &= (\mathbf{Y} - \mathbf{P}_X \mathbf{Y})^T \mathbf{1} \\ &= [(\mathbf{I} - \mathbf{P}_X) \mathbf{Y}]^T \mathbf{1} \end{aligned}$$

It is known that $\mathbf{I} - \mathbf{P}_X$ is the projection matrix onto the orthogonal complement of $\mathcal{C}(\mathbf{X})$, so $(\mathbf{I} - \mathbf{P}_X) \mathbf{Y}$ is in $\mathcal{C}(\mathbf{X})^\perp$. But since $\mathbf{1}$ is the first column of \mathbf{X} , $\mathbf{1} \in \mathcal{C}(\mathbf{X})$, so the inner product $\langle (\mathbf{I} - \mathbf{P}_X) \mathbf{Y}, \mathbf{1} \rangle = 0$.

-
- b. Let \mathbf{X}^* be a design matrix based on \mathbf{X} in which columns 2, ..., p are centered. That is, $X_{ij}^* = X_{ij} - \frac{1}{n} \sum_{i=1}^n X_{ij}$ for $j = 2, \dots, p$. Suppose we fit the model $\mathbf{Y} = \mathbf{X}^* \boldsymbol{\beta}^* + \boldsymbol{\epsilon}$. How does the value of $\hat{\boldsymbol{\beta}}^*$ compare to the value of $\hat{\boldsymbol{\beta}}$? (Hint start by partitioning $\mathbf{X} = [\mathbf{1}_n \quad \mathbf{W}]$. You may use the result $\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{F} \mathbf{E}^{-1} \mathbf{F}^T & -\mathbf{F} \mathbf{E}^{-1} \\ -\mathbf{E}^{-1} \mathbf{F}^T & \mathbf{E}^{-1} \end{bmatrix}$ where $\mathbf{E} = \mathbf{D} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B}$ and $\mathbf{F} = \mathbf{A}^{-1} \mathbf{B}$.)

Answer: Let \mathbf{W}^* denote the matrix of centered columns of \mathbf{X}^* . Then $\mathbf{X}^* = [\mathbf{1}_n \quad \mathbf{W}^*]$. Note that \mathbf{X}^* is still full rank, so we can get the unique projection matrix onto the column space of \mathbf{X}^* using the Moore-Penrose generalized inverse: $(\mathbf{X}^{*T} \mathbf{X}^*)^{-1}$. It should also be noted that $\frac{1}{n-1} \mathbf{W}^{*T} \mathbf{W}^*$ forms the variance-covariance matrix for columns 2 through p of \mathbf{X} . So let $(n-1) \boldsymbol{\Sigma}_X = \mathbf{W}^{*T} \mathbf{W}^*$.

In terms of \mathbf{X}^* , the LSE $\hat{\boldsymbol{\beta}}^*$ is

$$\begin{aligned} \hat{\boldsymbol{\beta}}^* &= (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \mathbf{Y} \\ &= \left(\begin{bmatrix} \mathbf{1} \\ \mathbf{W}^* \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{W}^* \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{1} \\ \mathbf{W}^* \end{bmatrix}^T \mathbf{Y} \\ &= \begin{bmatrix} \mathbf{1}^T \mathbf{1} & \mathbf{1}^T \mathbf{W}^* \\ \mathbf{W}^{*T} \mathbf{1} & \mathbf{W}^{*T} \mathbf{W}^* \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}^T \mathbf{Y} \\ \mathbf{W}^{*T} \mathbf{Y} \end{bmatrix} \\ &= \begin{bmatrix} n & \mathbf{0}_{1 \times (p-1)} \\ \mathbf{0}_{(p-1) \times 1} & (\mathbf{W}^{*T} \mathbf{W}^*)_{(p-1) \times (p-1)} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}^T \mathbf{Y} \\ \mathbf{W}^{*T} \mathbf{Y} \end{bmatrix} \end{aligned}$$

Using the result provided, we have

$$\begin{bmatrix} n & \mathbf{0}_{1 \times (p-1)} \\ \mathbf{0}_{(p-1) \times 1} & (\mathbf{W}^{*T} \mathbf{W}^*)_{(p-1) \times (p-1)} \end{bmatrix}^{-1} = \begin{bmatrix} 1/n & \mathbf{0}_{1 \times (p-1)} \\ \mathbf{0}_{(p-1) \times 1} & (\mathbf{W}^{*T} \mathbf{W}^*)^{-1} \end{bmatrix}$$

So,

$$\begin{aligned} \hat{\boldsymbol{\beta}}^* &= \begin{bmatrix} 1/n & \mathbf{0}_{1 \times (p-1)} \\ \mathbf{0}_{(p-1) \times 1} & (\mathbf{W}^{*T} \mathbf{W}^*)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1}^T \mathbf{Y} \\ \mathbf{W}^{*T} \mathbf{Y} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n} (\mathbf{1}^T \mathbf{Y}) + 0 \\ 0 + (\mathbf{W}^{*T} \mathbf{W}^*)^{-1} \mathbf{W}^{*T} \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \bar{Y} \\ (\mathbf{W}^{*T} \mathbf{W}^*)^{-1} \mathbf{W}^{*T} \mathbf{Y} \end{bmatrix} \end{aligned}$$

In contrast, the LSE for β when the columns of \mathbf{X} are not centered is

$$\hat{\beta} = \left[(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{Y} \right]$$

c. In the same setting as part (c), how does the variance of $\hat{\beta}$ compare to the variance of $\hat{\beta}^*$?

Answer: The variance of $\hat{\beta}^*$ is

$$\begin{aligned} \sigma^2 (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} &= \sigma^2 \left(\begin{bmatrix} \mathbf{1}^T \\ \mathbf{W}^{*T} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{W}^* \end{bmatrix} \right)^{-1} \\ &= \sigma^2 \begin{bmatrix} n & \mathbf{1}^T \mathbf{W}^* \\ \mathbf{W}^{*T} \mathbf{1} & \mathbf{W}^{*T} \mathbf{W}^* \end{bmatrix} \end{aligned}$$

Since \mathbf{X}^* is just a shifted version of \mathbf{X} , I feel like the variances would be the same.

Problem 3

Researchers wish to estimate two parameters θ and ϕ . It is possible to make three different kinds of observations: (a) observations with expectation $\theta + 3\phi$, (b) observations with expectation $\theta + \phi$, and (c) observations with expectation $\theta - 2\phi$. Suppose there are n_1 observations of type (a), n_2 observations of type (b), and n_3 observations of type (c). Assume the observations are all independent and have variance σ^2 .

- a. Write out the linear model corresponding to these observations.

Answer:

$$\begin{bmatrix} \mathbf{Y}_{1,(n_1 \times 1)} \\ \mathbf{Y}_{2,(n_2 \times 1)} \\ \mathbf{Y}_{3,(n_3 \times 1)} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{n_1} & 3 \cdot \mathbf{1}_{n_1} \\ \mathbf{1}_{n_2} & 1 \cdot \mathbf{1}_{n_2} \\ \mathbf{1}_{n_3} & -2 \cdot \mathbf{1}_{n_3} \end{bmatrix} \begin{bmatrix} \theta \\ \phi \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- b. For what choices of n_1 , n_2 , and n_3 will θ and ϕ be uncorrelated?

Answer: The variance-covariance matrix for $(\theta, \phi)^T$ is computed as

$$\sigma^2(\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 \begin{bmatrix} n & 3n_1 + n_2 - 2n_3 \\ 3n_1 + n_2 - 2n_3 & 9n_1 + n_2 + 4n_3 \end{bmatrix}^{-1}$$

The rank of \mathbf{X} is 2, so $\mathbf{X}^T \mathbf{X}$ is full rank and therefore nonsingular. The inverse of $\mathbf{X}^T \mathbf{X}$ requires that we switch the position of the diagonal terms and negate the off-diagonal terms, and divide all entries by the determinant of $\mathbf{X}^T \mathbf{X}$. After doing these steps, it is still the case that in order for θ and ϕ to be uncorrelated, the off-diagonal terms of $\mathbf{X}^T \mathbf{X}$, $3n_1 + n_2 - 2n_3$ must equal zero. This is the case if $n_1 = 2k_1$, $n_2 = 6k_2$, and $n_3 = 3k_3$, for non-negative, integer values of k_1, k_2 , and k_3 such that $k_1 + k_2 = k_3$.

Problem 4

For this question, use the subset of the `ChickWeight` dataset (part of base R). It contains data on the weight of chicks (`weight`, in grams), at specific ages (`Time`, in days) under different diets (`Diet`, factor with levels 1, 2, 3, 4). The variable `Chick` identifies unique chicks. Create a subset of this data that contains all measurements made at age 6 days. Use this dataset, which includes only one value per chick, for the remainder of this question.

a. Write out the linear model for chick weight as a function of diet. Make clear what the coding of the columns in your design matrix represent.

Answer:

$$Y = X\beta + \epsilon$$

where

$$Y_{49 \times 1} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{49} \end{bmatrix} \quad X_{49 \times 4} = \begin{bmatrix} 1_{19} & 0_{19} & 0_{19} & 0_{19} \\ 0_{10} & 1_{10} & 0_{10} & 0_{10} \\ 0_{10} & 0_{10} & 1_{10} & 0_{10} \\ 0_{10} & 0_{10} & 0_{10} & 1_{10} \end{bmatrix}, \quad \beta_{4 \times 1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, \quad \epsilon_{49 \times 1} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_{49} \end{bmatrix}$$

and

$$E[\epsilon] = 0$$

$$\text{Var}[\epsilon] = \sigma^2 \mathbf{I}_{49 \times 49}$$

In the design matrix X , the columns are coded in binary with 1s for the chicks on diet corresponding to the column number and 0s otherwise.

```
cw <- ChickWeight %>% # Keep only measurements for 6 days
  filter(Time == 6)

# Response [Weight]
Y <- cw$weight

# Design Matrix [Diet = 1, Diet = 2, Diet = 3, Diet = 4]
X <- cbind(as.numeric(cw$Diet == 1) , as.numeric(cw$Diet == 2)
           , as.numeric(cw$Diet == 3) , as.numeric(cw$Diet == 4)
           )

colnames(X) <- c("diet1" , "diet2"
                , "diet3", "diet4"
                )
```

b. What is the average weight of 6-day-old chicks for each diet? Provide your answer in the form of an estimable quantity $\mathbf{G}^T \boldsymbol{\beta}$.

Answer:

$$\mathbf{G}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y},$$

Where $\mathbf{G} = \mathbf{I}_{4 \times 4}$, so that the estimable quantity is reduced to

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \hat{\boldsymbol{\beta}} = \begin{bmatrix} \frac{1}{19} \sum_{i=1}^{19} Y_i \\ \frac{1}{10} \sum_{i=20}^{29} Y_i \\ \frac{1}{10} \sum_{i=30}^{39} Y_i \\ \frac{1}{10} \sum_{i=40}^{49} Y_i \end{bmatrix}.$$

In other words, an estimate for the average weight of 6-day-old chicks on each diet is the value of the corresponding estimated parameter, or the sample average for each group.

c. Find the corresponding BLUEs $\mathbf{G}^T \hat{\boldsymbol{\beta}}$. Report the point estimates and the estimated standard deviation (square root of estimated variance) of each estimate.

Answer: As stated in the previous part, BLUEs are the average weights of the chicks on each diet. For the point estimates, we can compute $\hat{\boldsymbol{\beta}}$ as $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ or simply find the averages of the weights for each diet:

```
beta_LSE <- solve(t(X) %*% X) %*% (t(X) %*% Y)
beta_LSE
```

```
##           [,1]
## diet1 66.78947
## diet2 75.40000
## diet3 77.90000
## diet4 83.90000
```

Since our BLUE was found under the assumption of constant variance of the errors, we can estimate the variance with the sample variance of all of the weights and multiply that by $(\mathbf{X}^T \mathbf{X})^{-1}$:

```
v <- var(Y)
beta_vcov <- v*solve(t(X) %*% X)
# Variance:
diag(beta_vcov)
```

```
##      diet1      diet2      diet3      diet4
## 4.274570 8.121684 8.121684 8.121684
```

```
# Standard Deviation
sqrt(diag(beta_vcov))
```

```
##      diet1      diet2      diet3      diet4
## 2.067503 2.849857 2.849857 2.849857
```

To Summarize:

Parameter	Estimate (g)	Standard Deviation (g)
β_1 (Avg. Chick Weight on Diet 1)	66.78947	2.067503
β_2 (Avg. Chick Weight on Diet 2)	75.40000	2.849857
β_3 (Avg. Chick Weight on Diet 3)	77.90000	2.849857
β_4 (Avg. Chick Weight on Diet 4)	83.90000	2.849857

Problem 5

Researchers wish to conduct a randomized clinical trial (RCT) to compare effect of a candidate drug on blood pressure in adults. They plan a simple study: n participants will be randomized to receive the candidate drug or the current standard of care (a competitor's drug) and at the end of follow-up will have their blood pressure measured. The goal of the trial is to estimate the difference in average blood pressure between the groups at the end of the trial.

a. Write out a linear model that can be used to answer the question of interest for this study. Let n_{txt} and n_{con} represent the number of participants randomized to the experimental drug ("treatment") and to the standard of care ("control"), so that $n_{txt} + n_{con} = n$. Identify what the data values and parameters in your model represent. (*You may wish to consider the remainder of this problem when determining how to parameterize your model*)

Answer: Instead of using n_{con} and n_{txt} , I'm going to use n_c and n_t , respectively.

The linear model that we will use to answer the question of interest is

$$\begin{bmatrix} \mathbf{Y}_c \\ \mathbf{Y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{n_c} & \mathbf{0}_{n_c} \\ \mathbf{0}_{n_t} & \mathbf{1}_{n_t} \end{bmatrix} \begin{bmatrix} \beta_c \\ \beta_t \end{bmatrix} + \boldsymbol{\epsilon}$$

Where \mathbf{Y}_c and \mathbf{Y}_t are the responses corresponding to those in the control and treatment group, respectively; the first column of the design matrix \mathbf{X} indicates the control group and the second column indicates the treatment group; β_c and β_t are the parameters representing the effect of no treatment and treatment, respectively, and $\boldsymbol{\epsilon}$ is the vector of random errors for each observation, which is assumed to have expectation $\mathbf{0}_n$ and variance-covariance matrix $\sigma^2 \mathbf{I}_{n \times n}$.

b. Find the BLUE in your model from (a) that answers the question of interest.

Answer: As was stated in the question, the goal is to estimate the difference in average blood pressure between the groups. This can be stated as estimating $\beta_t - \beta_c$ or $\mathbf{g}^T \boldsymbol{\beta}$, where $\mathbf{g}^T = [-1 \quad 1]$. This is estimable, so $\mathbf{g}^T \hat{\boldsymbol{\beta}}$ is the unique BLUE, by proposition 4.7. Computing this, we get

$$\begin{aligned} \mathbf{g}^T \hat{\boldsymbol{\beta}} &= \mathbf{g}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \\ &= [-1 \quad 1] \begin{bmatrix} 1/n_c & 0 \\ 0 & 1/n_t \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n_c} Y_i \\ \sum_{i=n_c+1}^n Y_i \end{bmatrix} \\ &= [-1 \quad 1] \begin{bmatrix} \bar{Y}_c \\ \bar{Y}_t \end{bmatrix} = \bar{Y}_t - \bar{Y}_c \end{aligned}$$

c. What choice(s) of n_{txt} and n_{con} minimizes the variance of the BLUE in (b)? Prove that your answer minimizes this variance.

Answer: The variance of $\mathbf{g}^T \hat{\boldsymbol{\beta}}$ is

$$\begin{aligned} \sigma^2 \mathbf{g}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{g} &= \sigma^2 [-1 \quad 1] \begin{bmatrix} 1/n_c & 0 \\ 0 & 1/n_t \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= \sigma^2 \left(\frac{1}{n_c} + \frac{1}{n_t} \right) \end{aligned}$$

To minimize the variance, we need to minimize $\frac{1}{n_c} + \frac{1}{n_t}$ along the constraint $n = n_c + n_t$. Using the constraint, we can make a substitution for n_t so that we have $\frac{1}{n_c} + \frac{1}{n - n_c}$. Taking the derivative with respect to n_c and

setting this equal to zero, we have $\frac{1}{(n-n_c)^2} - \frac{1}{n_c^2} = 0$. This implies that $n_c = n/2$ is a critical point. To show that it is a minimum, we can see that when $n_c < n/2$, $\frac{1}{(n-n_c)^2} - \frac{1}{n_c^2} < 0$ and $\frac{1}{(n-n_c)^2} - \frac{1}{n_c^2} > 0$ for $n_c > n/2$.

Problem 6

A different medical center has offered to partner with the company in Problem 3 for their RCT. The researchers decide to stratify the study by medical center, so that participants at each center are randomized to the two groups. Assume that the groups are equally split, so that:

- n_0 participants at center A receive the experimental drug
- n_0 participants at center A receive the control drug
- n_0 participants at center B receive the experimental drug
- n_0 participants at center B receive the control drug

Assume the populations at the centers are exchangeable, so that the effect of each drug is the same among participants at both centers. However, the instruments for measuring blood pressure at medical center B are older and so measurements made at medical center B have **twice the variance** of the measurements made at medical center A.

a. Write out a linear model that can be used to answer the question of interest for this study. Identify what the data values and parameters in your model represent.

Answer: A linear model that we can use to answer the question of interest (is there a difference between the treatment and control averages) is

$$\begin{bmatrix} \mathbf{Y}_{At} \\ \mathbf{Y}_{Ac} \\ \mathbf{Y}_{Bt} \\ \mathbf{Y}_{Bc} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{n_0} & \mathbf{0}_{n_0} & \mathbf{0}_{n_0} & \mathbf{0}_{n_0} \\ \mathbf{0}_{n_0} & \mathbf{1}_{n_0} & \mathbf{0}_{n_0} & \mathbf{0}_{n_0} \\ \mathbf{0}_{n_0} & \mathbf{0}_{n_0} & \mathbf{1}_{n_0} & \mathbf{0}_{n_0} \\ \mathbf{0}_{n_0} & \mathbf{0}_{n_0} & \mathbf{0}_{n_0} & \mathbf{1}_{n_0} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \boldsymbol{\epsilon}$$

Where \mathbf{Y}_{At} and \mathbf{Y}_{Ac} are the treatment and control group responses, respectively, at center A, and \mathbf{Y}_{Bt} and \mathbf{Y}_{Bc} are the treatment and control group responses, respectively at center B. The columns of the design matrix \mathbf{X} , from left to right, are indicators of the treatment group at center A, the control group at center A, the treatment group at center B and the control group at center B. β_1 and β_2 are the treatment and control effects, respectively. $\boldsymbol{\epsilon}$ is the vector of random errors for each observation, and it is assumed that

$$E[\boldsymbol{\epsilon}] = \mathbf{0}_n \quad \text{and} \quad \text{Var}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{V},$$

where

$$\mathbf{V} = \begin{bmatrix} \mathbf{I}_{2n_0 \times 2n_0} & \mathbf{0}_{2n_0 \times 2n_0} \\ \mathbf{0}_{2n_0 \times 2n_0} & 2\mathbf{I}_{2n_0 \times 2n_0} \end{bmatrix} \quad \text{and} \quad \mathbf{V}^{-1} = \begin{bmatrix} \mathbf{I}_{2n_0 \times 2n_0} & \mathbf{0}_{2n_0 \times 2n_0} \\ \mathbf{0}_{2n_0 \times 2n_0} & \frac{1}{2}\mathbf{I}_{2n_0 \times 2n_0} \end{bmatrix}$$

b. Find the BLUE from your model from (a) that answers the question of interest.

Answer: Because the variance of the errors are not constant, we should use generalized least squares to find the BLUE that answers the question of whether there is a difference between the treatment and control averages. Here, our BLUE will be $\mathbf{g}^T \hat{\boldsymbol{\beta}}^*$, where $\hat{\boldsymbol{\beta}}^* = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}$. So we have

$$\begin{aligned} \hat{\boldsymbol{\beta}}^* &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} \\ &= \begin{bmatrix} n_0 & 0 & 0 & 0 \\ 0 & n_0 & 0 & 0 \\ 0 & 0 & \frac{n_0}{2} & 0 \\ 0 & 0 & 0 & \frac{n_0}{2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_{n_0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{2}\mathbf{1}_{n_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{2}\mathbf{1}_{n_0} \end{bmatrix}^T \begin{bmatrix} \mathbf{Y}_{At} \\ \mathbf{Y}_{Ac} \\ \mathbf{Y}_{Bt} \\ \mathbf{Y}_{Bc} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n_0} & 0 & 0 & 0 \\ 0 & \frac{1}{n_0} & 0 & 0 \\ 0 & 0 & \frac{2}{n_0} & 0 \\ 0 & 0 & 0 & \frac{2}{n_0} \end{bmatrix} \begin{bmatrix} \sum_{At} Y_i \\ \sum_{Ac} Y_i \\ \sum_{Bt} Y_i \\ \sum_{Bc} Y_i \end{bmatrix} = \begin{bmatrix} \bar{Y}_{At} \\ \bar{Y}_{Ac} \\ \bar{Y}_{Bt} \\ \bar{Y}_{Bc} \end{bmatrix} \end{aligned}$$

And the BLUE will be $\mathbf{g}^T \hat{\boldsymbol{\beta}}^*$ where $\mathbf{g}^T = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$:

$$\mathbf{g}^T \hat{\boldsymbol{\beta}}^* = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \bar{Y}_{At} \\ \bar{Y}_{Ac} \\ \bar{Y}_{Bt} \\ \bar{Y}_{Bc} \end{bmatrix} = \frac{(\bar{Y}_{At} + \bar{Y}_{Bt})}{2} - \frac{(\bar{Y}_{Ac} + \bar{Y}_{Bc})}{2}$$
