

# MAST30025: Linear Statistical Models

## Solutions to Weeks 12 and 13 Lab

1. The following data were collected, to compare two treatments. The treatments were randomly assigned to test subjects.

treatment 1		treatment 2	
subject	response	subject	response
10	7.5	11	9.5
9	9.6	6	9.7
5	8.4	2	10.8
12	10.6	8	11.9
7	9.9	4	10.0
1	10.6	3	12.9

- (a) Estimate the difference between treatment effects, and test if it is significantly different from 0.

**Solution:**

```
> mydata <- data.frame(response = c(7.5, 9.6, 8.4, 10.6, 9.9, 10.6,
+                                9.5, 9.7, 10.8, 11.9, 10.0, 12.9),
+                        treatment = factor(rep(c(1,2), c(6,6))))
> summary(lm(response ~ treatment, mydata))
```

Call:

```
lm(formula = response ~ treatment, data = mydata)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.93333	-1.05000	0.08333	1.11667	2.10000

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.4333	0.5312	17.757	6.84e-09 ***
treatment2	1.3667	0.7513	1.819	0.0989 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.301 on 10 degrees of freedom

Multiple R-squared: 0.2486, Adjusted R-squared: 0.1735

F-statistic: 3.309 on 1 and 10 DF, p-value: 0.09893

The estimated difference between treatment effects is 1.3667. The  $p$ -value for the test of  $\tau_2 - \tau_1 = 0$  (against a general alternative) is 0.0989, which is borderline significant (but insignificant at a 5% level).

- (b) Now suppose that it is discovered that the response can be affected by the season, and that the data was collected over a period of six months, in the order given by the table. That is, a month was spent collecting each row of the table.

We re-express the experiment by blocking: each month (row of the table) is considered one block, and we model the data as an additive two-factor model (the factors being the treatment and the block). Using this model, repeat your analysis. Is the estimate different? Is the  $p$ -value different?

**Solution:**

```
> mydata$block <- factor(rep(1:6, 2))
> model <- lm(response ~ treatment + block, mydata)
> summary(model)
```

```

Call:
lm(formula = response ~ treatment + block, data = mydata)

Residuals:
    1     2     3     4     5     6     7     8 
-0.31667  0.63333 -0.51667  0.03333  0.63333 -0.46667  0.31667 -0.63333 
    9    10    11    12 
 0.51667 -0.03333 -0.63333  0.46667 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.8167      0.5692  13.734 3.67e-05 ***
treatment2    1.3667      0.4302   3.176  0.02464 *
block2         1.1500      0.7452   1.543  0.18343
block3         1.1000      0.7452   1.476  0.19994
block4         2.7500      0.7452   3.690  0.01414 *
block5         1.4500      0.7452   1.946  0.10926
block6         3.2500      0.7452   4.361  0.00728 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7452 on 5 degrees of freedom
Multiple R-squared:  0.8768,    Adjusted R-squared:  0.7289 
F-statistic:  5.93 on 6 and 5 DF,  p-value: 0.03493

> anova(model)

Analysis of Variance Table

Response: response
      Df Sum Sq Mean Sq F value Pr(>F)
treatment  1  5.6033   5.6033 10.0900 0.02464 *
block      5 14.1567   2.8313  5.0984 0.04910 *
Residuals  5  2.7767   0.5553
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The estimated difference between treatment effects is still 1.3667. It has not changed because the new design is a complete block design. The  $p$ -value for the test of  $\tau_2 - \tau_1 = 0$  (against a general alternative) is now 0.0246, which is significant. The blocking has reduced the variability of our estimate and thus increased the power of our test.

2. In lectures, we showed that for the (randomised) complete block design

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij},$$

a solution to the reduced normal equations for  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_k)^T$  is given by

$$(\bar{y}_{\cdot 1} - \bar{y}_{\cdot \cdot}, \dots, \bar{y}_{\cdot k} - \bar{y}_{\cdot \cdot})^T.$$

Here we suppose that we have  $b$  blocks and  $k$  treatments.

Consider now the completely randomised design, with  $k$  treatments and  $b$  replications of each treatment

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}.$$

Treating  $\mu$  as a nuisance parameter, obtain the reduced normal equations for  $\boldsymbol{\tau}$ , then show that they admit the solution

$$(\bar{y}_1 - \bar{y}_{\cdot}, \dots, \bar{y}_k - \bar{y}_{\cdot})^T.$$

**Solution:** We have  $\mathbf{y} = \mathbf{1}_n \mu + X_2 \boldsymbol{\tau} + \boldsymbol{\varepsilon}$ , that is  $X_1 = \mathbf{1}_n$ .

Thus  $X_1^T X_1 = n = kb$ ,  $(X_1^T X_1)^c = 1/n$ ,  $H_1 = \frac{1}{n}J_n$ , and  $X_{2|1} = (I_n - \frac{1}{n}J_n)X_2 = X_2 - \frac{1}{k}J_{n,k}$ , where  $J_{n,k}$  is the  $n \times k$  matrix of ones.

The reduced normal equations are  $X_{2|1}^T X_{2|1} \mathbf{t} = X_{2|1}^T \mathbf{y}$ , that is

$$b(I_k - \frac{1}{k}J_k)\mathbf{t} = X_2^T(I_n - \frac{1}{n}J)\mathbf{y}.$$

Since  $[b(I_k - \frac{1}{k}J_k)]^c = \frac{1}{b}I_k$ , we have

$$\mathbf{t} = \frac{1}{b}X_2^T(I_n - \frac{1}{n}J_n)\mathbf{y} = \frac{1}{b}(X_2^T - \frac{1}{k}J_{k,n})\mathbf{y} = \begin{bmatrix} \bar{y}_1 - \bar{y} \\ \vdots \\ \bar{y}_k - \bar{y} \end{bmatrix},$$

noting that

$$X_2 = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 0 \\ \hline 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \dots & 0 \\ \hline & & \ddots & \\ \hline 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Subject to relabelling the indices, this is exactly the same solution as for the reduced normal equations for the CBD.

3. Suppose we have a (randomised) complete block design,  $y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$ , with  $b$  blocks and  $k$  treatments.

Let  $\mathbf{c}^T$  be a treatment contrast, so that  $\mathbf{c}^T \boldsymbol{\tau}$  is estimable, in which case

$$\text{Var } \mathbf{c}^T \boldsymbol{\tau} = \frac{\sigma^2}{b} \sum_{j=1}^k c_j^2.$$

- (a) Give an approximate  $100(1-\alpha)\%$  CI for  $\mathbf{c}^T \boldsymbol{\tau}$ , using the percentage point from a normal rather than the correct percentage point from a  $t$  distribution. (This is reasonable if the degrees of freedom are large.)

**Solution:** Let  $s^2$  be the sample variance, then an approximate CI is

$$\sum_j c_j \bar{y}_{.j} \pm z_{\alpha/2} s \sqrt{\sum c_j^2 / b}.$$

- (b) Now suppose that you know  $\sigma^2$  (perhaps you have an estimate from a pilot study), and that you think a plausible alternative to  $\mathbf{c}^T \boldsymbol{\tau} = 0$  is given by some  $\mathbf{c}^T \boldsymbol{\tau}^* \neq 0$ . How large should  $b$  be to give a power of  $100(1-\alpha)\%$  against this alternative (roughly)?

**Solution:** To have power  $100(1-\alpha)\%$  against this alternative we need  $|\mathbf{c}^T \boldsymbol{\tau}^*| > z_{\alpha/2} \sigma \sqrt{\sum c_j^2 / b}$ , which gives

$$b > z_{\alpha/2}^2 \sigma^2 \sum_j c_j^2 / (\mathbf{c}^T \boldsymbol{\tau}^*)^2 = z_{\alpha/2}^2 \sigma^2 \mathbf{c}^T \mathbf{c} / (\mathbf{c}^T \boldsymbol{\tau}^*)^2.$$

4. Consider the following data:

Response	Block	Treatment
1.245	1	1
1.804	1	2
2.468	2	1
6.664	2	3
5.573	3	1
-0.560	3	4
7.880	4	2
10.469	4	3
0.457	5	2
-3.621	5	4
-4.291	6	3
-9.384	6	4

(a) Show that this data comes from a balanced incomplete block design, and give  $t$ ,  $b$ ,  $k$ ,  $r$  and  $\lambda$ .

**Solution:** Each treatment occurs at most once in a block, and in exactly 3 blocks. Each pair of treatments occurs in exactly 1 block. Thus we have a BIBD with  $t = 4$ ,  $b = 6$ ,  $k = 2$ ,  $r = 3$  and  $\lambda = 1$ .

(b) Give the design matrix  $X^A$  for a model with block and treatment effects (and an overall mean).

**Solution:** Taking the responses in the order given, and  $\beta = (\mu, \beta_1, \dots, \beta_6, \tau_1, \dots, \tau_4)^T$ , where the  $\beta_i$  are the block effects and the  $\tau_j$  are the treatment effects, we get the design matrix

```
> XA <- matrix(0, nrow = 12, ncol = 11)
> XA[,1] <- 1
> XA[cbind(1:12, rep(1:6, each=2)+1)] <- 1
> XA[cbind(1:12, c(1,2,1,3,1,4,2,3,2,4,3,4)+7)] <- 1
> XA
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]
[1,]	1	1	0	0	0	0	0	1	0	0	0
[2,]	1	1	0	0	0	0	0	0	1	0	0
[3,]	1	0	1	0	0	0	0	1	0	0	0
[4,]	1	0	1	0	0	0	0	0	0	1	0
[5,]	1	0	0	1	0	0	0	1	0	0	0
[6,]	1	0	0	1	0	0	0	0	0	0	1
[7,]	1	0	0	0	1	0	0	0	1	0	0
[8,]	1	0	0	0	1	0	0	0	0	1	0
[9,]	1	0	0	0	0	1	0	0	1	0	0
[10,]	1	0	0	0	0	1	0	0	0	0	1
[11,]	1	0	0	0	0	0	1	0	0	1	0
[12,]	1	0	0	0	0	0	1	0	0	0	1

(c) Using this model, estimate  $\tau_1 - \tau_2$ , the difference between the first two treatment effects, and its variance. Write the variance estimate as  $s^2 \mathbf{c}^T (X^{AT} X^A)^c \mathbf{c}$  for a suitable  $\mathbf{c}$ .

**Solution:**

```
> library(MASS)
> library(Matrix)
> y <- c(1.245, 1.804, 2.468, 6.664, 5.573, -0.560,
+       7.880, 10.469, 0.457, -3.621, -4.291, -9.384)
> bA <- ginv(t(XA) %*% XA) %*% t(XA) %*% y
> cA <- c(0, 0, 0, 0, 0, 0, 0, 1, -1, 0, 0)
> n <- 12
> (rA <- rankMatrix(XA)[1])

[1] 9
> (s2A <- sum((y - XA %*% bA)^2)/(n-rA))
```

```

[1] 1.57602
> t(cA) %*% bA
      [,1]
[1,] -0.1675
> t(cA) %*% ginv(t(XA) %*% XA) %*% cA
      [,1]
[1,] 1
> s2A * t(cA) %*% ginv(t(XA) %*% XA) %*% cA
      [,1]
[1,] 1.57602

```

- (d) Give the design matrix  $X^B$  for a model with just treatment effects (and an overall mean).

**Solution:**

```

> XB <- matrix(0, nrow = 12, ncol = 5)
> XB[,1] <- 1
> XB[cbind(1:12,c(1,2,1,3,1,4,2,3,2,4,3,4)+1)] <- 1
> XB
      [,1] [,2] [,3] [,4] [,5]
[1,] 1 1 0 0 0
[2,] 1 0 1 0 0
[3,] 1 1 0 0 0
[4,] 1 0 0 1 0
[5,] 1 1 0 0 0
[6,] 1 0 0 0 1
[7,] 1 0 1 0 0
[8,] 1 0 0 1 0
[9,] 1 0 1 0 0
[10,] 1 0 0 0 1
[11,] 1 0 0 1 0
[12,] 1 0 0 0 1

```

- (e) Using this model, estimate  $\tau_1 - \tau_2$ , the difference between the first two treatment effects, and its variance. Write the variance estimate as  $s^2 \mathbf{c}^T (X^{BT} X^B)^c \mathbf{c}$  for a suitable  $\mathbf{c}$ .

**Solution:**

```

> bB <- ginv(t(XB) %*% XB) %*% t(XB) %*% y
> cB <- c(0, 1, -1, 0, 0)
> (rB <- rankMatrix(XB)[1])
[1] 4
> (s2B <- sum((y - XB %*% bB)^2)/(n-rB))
[1] 24.85392
> t(cB) %*% bB
      [,1]
[1,] -0.285
> t(cB) %*% ginv(t(XB) %*% XB) %*% cB
      [,1]
[1,] 0.6666667
> s2B * t(cB) %*% ginv(t(XB) %*% XB) %*% cB
      [,1]
[1,] 16.56928

```

- (f) Show that when going from model A (BIBD) to model B (CRD) the term  $\mathbf{c}^T (X^T X)^c \mathbf{c}$  decreases, but  $s^2$  increases markedly. What does this indicate?

**Solution:** See above. This indicates that the blocks are effective in their intended purpose (reducing variance).

(g) Is your estimate for  $\tau_1 - \tau_2$  the same or different for the two models? Why?

**Solution:** The estimates are different. This is because in the BIBD, the blocks are not orthogonal to the treatments.

5. Consider the BIBD model, with  $t$  treatments and  $b$  blocks of size  $k$ . Let  $\lambda$  be the number of times each pair appears, and write the design as

$$\mathbf{y} = X_1\boldsymbol{\alpha} + X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}.$$

Show that for this model, contrasts in  $\boldsymbol{\tau}$  are estimable.

If  $\mathbf{c}^T\boldsymbol{\tau}$  is a contrast, show that an unbiased estimate is  $(k/\lambda t)\mathbf{c}^T\mathbf{q}$ , where

$$\mathbf{q} = \mathbf{t} - X_2^T X_1 \mathbf{b},$$

and  $\mathbf{t}$  are the treatment totals and  $\mathbf{b}$  the block totals.

**Solution:** From the lectures, we have that

$$\begin{aligned} X_{2|1}^T X_{2|1} &= \frac{\lambda t}{k} \left( I_t - \frac{1}{t} J_t \right) \\ (X_{2|1}^T X_{2|1})^c &= \frac{k}{\lambda t} I_t \end{aligned}$$

and therefore for any contrast  $\mathbf{c}^T\boldsymbol{\tau}$ ,

$$\begin{aligned} \mathbf{c}^T (X_{2|1}^T X_{2|1})^c X_{2|1}^T X_{2|1} &= \mathbf{c}^T \left( I_t - \frac{1}{t} J_t \right) \\ &= \mathbf{c}^T - \frac{1}{t} \mathbf{c}^T J_t \\ &= \mathbf{c}^T. \end{aligned}$$

The estimate follows from the solution to the reduced normal equations.

6. An experimenter is tasked with designing an experiment to compare three treatment levels. There is a known confounding factor, so a blocked design is appropriate. Consider the following two designs, each using four blocks of size three:

Design 1	block 1:	1	2	3
	block 2:	2	1	3
	block 3:	1	3	2
	block 4:	3	2	1
Design 2	block 1:	1	1	2
	block 2:	2	2	3
	block 3:	3	3	1
	block 4:	1	2	3

- (a) Which design is a complete block design?

**Solution:** The first design is a complete block design: each treatment appears exactly once in each block.

- (b) Write down the design matrix for each design. Hence show that  $\tau_2 - \tau_1$  is estimable in each case.

**Solution:**

```
> Xa <- matrix(0, 12, 8)
> Xa[,1] <- 1
> Xa[1:3,2] <- 1
> Xa[4:6,3] <- 1
> Xa[7:9,4] <- 1
> Xa[10:12,5] <- 1
> Xa[,6] <- c(1,0,0, 0,1,0, 1,0,0, 0,0,1)
```

```

> Xa[,7] <- c(0,1,0, 1,0,0, 0,0,1, 0,1,0)
> Xa[,8] <- c(0,0,1, 0,0,1, 0,1,0, 1,0,0)
> Xb <- Xa
> Xb[,6] <- c(1,1,0, 0,0,0, 0,0,1, 1,0,0)
> Xb[,7] <- c(0,0,1, 1,1,0, 0,0,0, 0,1,0)
> Xb[,8] <- c(0,0,0, 0,0,1, 1,1,0, 0,0,1)
> Xa

```

```

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    1    1    0    0    0    1    0    0
[2,]    1    1    0    0    0    0    1    0
[3,]    1    1    0    0    0    0    0    1
[4,]    1    0    1    0    0    0    1    0
[5,]    1    0    1    0    0    1    0    0
[6,]    1    0    1    0    0    0    0    1
[7,]    1    0    0    1    0    1    0    0
[8,]    1    0    0    1    0    0    0    1
[9,]    1    0    0    1    0    0    1    0
[10,]   1    0    0    0    1    0    0    1
[11,]   1    0    0    0    1    0    1    0
[12,]   1    0    0    0    1    1    0    0

```

```

> Xb

```

```

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    1    1    0    0    0    1    0    0
[2,]    1    1    0    0    0    1    0    0
[3,]    1    1    0    0    0    0    1    0
[4,]    1    0    1    0    0    0    1    0
[5,]    1    0    1    0    0    0    1    0
[6,]    1    0    1    0    0    0    0    1
[7,]    1    0    0    1    0    0    0    1
[8,]    1    0    0    1    0    0    0    1
[9,]    1    0    0    1    0    1    0    0
[10,]   1    0    0    0    1    1    0    0
[11,]   1    0    0    0    1    0    1    0
[12,]   1    0    0    0    1    0    0    1

```

```

> library(MASS)

```

```

> t <- c(0,0,0,0,0,-1,1,0)

```

```

> round(t(t) %*% ginv(t(Xa) %*% Xa) %*% t(Xa) %*% Xa, 10)

```

```

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    0    0    0    0   -1    1    0

```

```

> round(t(t) %*% ginv(t(Xb) %*% Xb) %*% t(Xb) %*% Xb, 10)

```

```

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    0    0    0    0   -1    1    0

```

- (c) For each design, in terms of the unknown error variance  $\sigma^2$ , what is the variance of the estimator for  $\tau_2 - \tau_1$ , the difference between the first two treatment effects?

Based on this, which design is better?

**Solution:**

```

> t(t) %*% ginv(t(Xa) %*% Xa) %*% t

```

```

      [,1]
[1,]  0.5

```

```

> t(t) %*% ginv(t(Xb) %*% Xb) %*% t

```

```

      [,1]
[1,] 0.6666667

```

The variance of the estimator is  $(1/2)\sigma^2$  for the first design and  $(2/3)\sigma^2$  for the second design. Clearly we prefer the CBD, as it gives the smaller variance.

7. In some situations, it is sensible to think of block effects as random. For example, experiments performed on a single day might be considered as a single block, subject to some effect for conditions on that day.

Consider the following model for an experiment with fixed treatment effects  $\boldsymbol{\tau}$  and random block effects  $\boldsymbol{\beta}$  (independent of the error  $\boldsymbol{\varepsilon}$ ):

$$\mathbf{y} = X_1\boldsymbol{\beta} + X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}, \quad \mathbb{E}\boldsymbol{\varepsilon} = \mathbf{0}, \quad \text{Var } \boldsymbol{\varepsilon} = \sigma^2 I, \quad \mathbb{E}\boldsymbol{\beta} = \mu\mathbf{1}, \quad \text{Var } \boldsymbol{\beta} = \sigma_\beta^2 I.$$

- (a) Find  $\mathbb{E}\mathbf{y}$  and  $V = \text{Var } \mathbf{y}$ .

**Solution:**  $\mathbb{E}\mathbf{y} = \mu X_1\mathbf{1} + X_2\boldsymbol{\tau} = \mu\mathbf{1} + X_2\boldsymbol{\tau}$  and  $\text{Var } \mathbf{y} = \sigma_\beta^2 X_1 X_1^T + \sigma^2 I$ .

- (b) Give a solution to the generalised least squares problem:

$$\min_{\mathbf{t}} (\mathbf{y} - X_2\mathbf{t})^T V^{-1} (\mathbf{y} - X_2\mathbf{t}).$$

**Solution:** Differentiating and setting the derivative to zero, we obtain the generalised normal equations

$$X_2^T V^{-1} X_2 \mathbf{t} = X_2^T V^{-1} \mathbf{y}.$$

Thus a solution is  $\mathbf{t} = (X_2^T V^{-1} X_2)^c X_2^T V^{-1} \mathbf{y}$ .

- (c) A problem with the generalised least squares above is that  $\mu$  may not be zero, so that if we write  $\mathbf{y} = X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}'$ , then  $\boldsymbol{\varepsilon}' = \boldsymbol{\varepsilon} + X_1\boldsymbol{\beta}$  does not have a zero mean.

To get around this, first suppose that each block is of size  $k$ , so

$$X_1 = \begin{bmatrix} \mathbf{1}_k & \mathbf{0}_k & \cdots & \mathbf{0}_k \\ \mathbf{0}_k & \mathbf{1}_k & \cdots & \mathbf{0}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_k & \mathbf{0}_k & \cdots & \mathbf{1}_k \end{bmatrix},$$

then suppose that  $U$  is such that  $X_1^T U = \mathbf{0}$ .

Put  $\mathbf{y}_1 = U^T \mathbf{y}$  and  $\mathbf{y}_2 = X_1^T \mathbf{y}$ , then show that we can write them as linear models whose errors have mean zero.

**Solution:** We have

$$\begin{aligned} \mathbf{y}_1 &= U^T X_2 \boldsymbol{\tau} + \boldsymbol{\varepsilon}_1, \quad \mathbb{E}\boldsymbol{\varepsilon}_1 = \mathbf{0}, \quad \text{Var } \boldsymbol{\varepsilon}_1 = \sigma^2 U^T U, \\ \mathbf{y}_2 &= k\mu\mathbf{1} + X_1^T X_2 \boldsymbol{\tau} + \boldsymbol{\varepsilon}_2, \quad \mathbb{E}\boldsymbol{\varepsilon}_2 = \mathbf{0}, \quad \text{Var } \boldsymbol{\varepsilon}_2 = (k^2\sigma_\beta^2 + k\sigma^2)I. \end{aligned}$$

- (d) Show that  $\text{Cov}(\mathbf{y}_1, \mathbf{y}_2) = \mathbb{E}(\mathbf{y}_1 - \mathbb{E}\mathbf{y}_1)(\mathbf{y}_2 - \mathbb{E}\mathbf{y}_2)^T = \mathbf{0}$ .

**Solution:** We have

$$\begin{aligned} \text{Cov}(\mathbf{y}_1, \mathbf{y}_2) &= \mathbb{E}U^T(\mathbf{y} - \mathbb{E}\mathbf{y})(\mathbf{y} - \mathbb{E}\mathbf{y})^T X_1 \\ &= U^T \text{Var } \mathbf{y} X_1 \\ &= U^T (\sigma^2 I + \sigma_\beta^2 X_1 X_1^T) X_1 \\ &= \sigma^2 U^T X_1 + \sigma_\beta^2 U^T X_1 X_1^T X_1 \\ &= \mathbf{0}. \end{aligned}$$