Consider the random blocks model $y_{i,j} = \mu + \alpha_i + \beta_j + \epsilon_{i,j}$ with $\epsilon_{i,j} \sim \mathsf{N}(0, \sigma^2)$, $i = 1, \ldots, 3$ and $j = 1, \ldots, 3$.

1. (15 pts) Write this model in matrix form as $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\theta} = (\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3)$.

   Solution:

$$\mathbf{y} = \begin{pmatrix} y_{1,1} \\ y_{1,2} \\ y_{1,3} \\ y_{2,1} \\ y_{2,2} \\ y_{2,3} \\ y_{3,1} \\ y_{3,2} \\ y_{3,3} \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \qquad \boldsymbol{\epsilon} \sim \mathsf{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

   ◇

2. (10 pts) What is the rank of the design matrix $\mathbf{X}$? Justify your answer.

   Solution: The rank is 5 (the fourth column of $\mathbf{X}$ is linearly dependent of the first 3, while the last column is linearly dependent of the first, fifth and sixth). ◇

3. (35 pts) List one possible set of constraints on $\boldsymbol{\theta}$ that makes the corresponding constrained least squares estimator be unique, and provide the solution to the normal equations under those constraints. (Hint: Rather than solving the normal equations, try to propose a solution and show that it satisfies the normal equations.)

   Solution: The simplest set of constraints is $\alpha_1 = 0$ and $\beta_1 = 0$. However, a simpler set of constraints to work with are $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$.

   The normal equations in this case correspond to $\mathbf{X}^T\mathbf{X} = \mathbf{X}^T\mathbf{y}$ where

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 9 & 3 & 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 0 & 0 & 1 & 1 & 1 \\ 3 & 0 & 3 & 0 & 1 & 1 & 1 \\ 3 & 0 & 0 & 3 & 1 & 1 & 1 \\ 3 & 1 & 1 & 1 & 3 & 0 & 0 \\ 3 & 1 & 1 & 1 & 0 & 3 & 0 \\ 3 & 1 & 1 & 1 & 0 & 0 & 3 \end{pmatrix} \qquad \mathbf{X}^T\mathbf{y} = \begin{pmatrix} y_{\cdot,\cdot} \\ y_{1,\cdot} \\ y_{2,\cdot} \\ y_{3,\cdot} \\ y_{\cdot,1} \\ y_{\cdot,2} \\ y_{\cdot,4} \end{pmatrix}$$

Because of the interpretation of the coefficients, it is natural to assume that the solution is of the form

$$\begin{pmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} \bar{y}_{\cdot\cdot} \\ \bar{y}_{1,\cdot} - \bar{y}_{\cdot\cdot} \\ \bar{y}_{2,\cdot} - \bar{y}_{\cdot\cdot} \\ \bar{y}_{3,\cdot} - \bar{y}_{\cdot\cdot} \\ \bar{y}_{\cdot,1} - \bar{y}_{\cdot\cdot} \\ \bar{y}_{\cdot,2} - \bar{y}_{\cdot\cdot} \\ \bar{y}_{\cdot,3} - \bar{y}_{\cdot\cdot} \end{pmatrix}$$

which indeed satisfy the normal equations. For example, for the first row of the normal equations,

$$\begin{aligned} 9\bar{y}_{\cdot\cdot} + 3(\bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot}) + 3(\bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot}) + 3(\bar{y}_{3\cdot} - \bar{y}_{\cdot\cdot}) + 3(\bar{y}_{\cdot 1} - \bar{y}_{\cdot\cdot}) + 3(\bar{y}_{\cdot 2} - \bar{y}_{\cdot\cdot}) + 3(\bar{y}_{\cdot 3} - \bar{y}_{\cdot\cdot}) \\ = -y_{\cdot\cdot} + y_{1\cdot} + y_{2\cdot} + y_{3\cdot} + y_{\cdot 1} + y_{\cdot 2} + y_{\cdot 3} \\ = -y_{\cdot\cdot} + 2y_{\cdot\cdot} = y_{\cdot\cdot} \end{aligned}$$

while for the second row

$$\begin{aligned} &3\bar{y}_{\cdot\cdot} + 3(\bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot}) + (\bar{y}_{\cdot 1} - \bar{y}_{\cdot\cdot}) + (\bar{y}_{\cdot 2} - \bar{y}_{\cdot\cdot}) + (\bar{y}_{\cdot 3} - \bar{y}_{\cdot\cdot}) \\ &\quad = -3\bar{y}_{\cdot\cdot} + 3\bar{y}_{1\cdot} + \bar{y}_{\cdot 1} + \bar{y}_{\cdot 2} + \bar{y}_{\cdot 3} = -\frac{1}{3}y_{\cdot\cdot} + y_{1\cdot} + \frac{1}{3}(y_{\cdot 1} + y_{\cdot 2} + y_{\cdot 3}) = y_{1\cdot} \end{aligned}$$

(the remaining of the rows are analogous to the second). $\diamond$

4. Suppose that we are interested in testing the hypotheses $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ against the alternative H_a : At least one $\alpha_i \neq 0$.

   (a) (10 pts) Write these hypotheses as a general linear hypotheses $H_0 : \mathbf{K}^T\boldsymbol{\theta} = \mathbf{m}$ and $H_a : \mathbf{K}^T\boldsymbol{\theta} \neq \mathbf{m}$. Make sure to show that $\mathbf{K}^T\boldsymbol{\theta} = \mathbf{m}$ is testable!

   Solution: Let

$$\mathbf{K}^T = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{m} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

   Note that the columns of $\mathbf{K}$ are linearly independent and that each of the rows of $\mathbf{K}^T\boldsymbol{\theta}$ is estimable. This is not the only possible configuration, another option is

$$\mathbf{K}^T = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{m} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

$\diamond$

(b) (30 pts) Describe a test (i.e., a statistic, its distribution under the null and the alternative, and a rejection region) for the general linear hypotheses described before.

Solution: Let's focus first on the statistic. The general result is $\mathbf{K}^T\hat{\boldsymbol{\theta}} \sim \mathsf{N}\left(\mathbf{K}^T\boldsymbol{\theta}, \sigma^2\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{K}\right)$. However, working with the full form is cumbersome, as it involves a generalized inverse of a large matrix. Instead, note that

$$\mathbf{K}^T\hat{\boldsymbol{\theta}} = \begin{pmatrix} \bar{y}_{1\cdot} - \bar{y}_{2\cdot} \\ \bar{y}_{1\cdot} - \bar{y}_{3\cdot} \end{pmatrix}$$

Since observations are independent we have that

$$\begin{pmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{3\cdot} \end{pmatrix} \sim \mathsf{N}\left(\begin{pmatrix} \mu + \alpha_1 + \frac{1}{3}\{\beta_1 + \beta_2 + \beta_3\} \\ \mu + \alpha_2 + \frac{1}{3}\{\beta_1 + \beta_2 + \beta_3\} \\ \mu + \alpha_3 + \frac{1}{3}\{\beta_1 + \beta_2 + \beta_3\} \end{pmatrix}, \frac{\sigma^2}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right),$$

and

$$\begin{pmatrix} \bar{y}_{1\cdot} - \bar{y}_{2\cdot} \\ \bar{y}_{1\cdot} - \bar{y}_{3\cdot} \end{pmatrix} \sim \mathsf{N}\left(\begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_1 - \alpha_3 \end{pmatrix}, \frac{\sigma^2}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \right).$$

So that

$$\begin{aligned} SSR &= \frac{1}{\sigma^2}\begin{pmatrix} \bar{y}_{1\cdot} - \bar{y}_{2\cdot} \\ \bar{y}_{1\cdot} - \bar{y}_{3\cdot} \end{pmatrix}^T \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \bar{y}_{1\cdot} - \bar{y}_{2\cdot} \\ \bar{y}_{1\cdot} - \bar{y}_{3\cdot} \end{pmatrix} \\ &= \frac{2}{\sigma^2}\left\{ (\bar{y}_{1\cdot} - \bar{y}_{2\cdot})^2 - (\bar{y}_{1\cdot} - \bar{y}_{2\cdot})(\bar{y}_{1\cdot} - \bar{y}_{3\cdot}) + (\bar{y}_{1\cdot} - \bar{y}_{3\cdot})^2 \right\} \sim \chi^2_2(\eta), \end{aligned}$$

where

$$\eta = \frac{1}{2}\begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_1 - \alpha_3 \end{pmatrix}^T \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_1 - \alpha_3 \end{pmatrix}.$$

On the other hand, the unbiased estimate of the variance σ^2 is

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^3 \sum_{j=1}^3 (y_{i,j} - \hat{y}_{i,j})^2}{4} = \frac{\sum_{i=1}^3 \sum_{j=1}^3 (y_{i,j} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2}{4},$$

which is independent of SSR and, after appropriate renormalization, follows a χ^2_4 distribution. Hence, the statistic

$$F = \frac{2\left\{ (\bar{y}_{1\cdot} - \bar{y}_{2\cdot})^2 - (\bar{y}_{1\cdot} - \bar{y}_{2\cdot})(\bar{y}_{1\cdot} - \bar{y}_{3\cdot}) + (\bar{y}_{1\cdot} - \bar{y}_{3\cdot})^2 \right\}}{\sum_{i=1}^3 \sum_{j=1}^3 (y_{i,j} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2} \frac{4}{2}$$

follows a non central F distribution, $F_{2,4}(\eta)$. Under the null hypotheses, $\eta = 0$ and an appropriate rejection region can be obtained by finding c such that $Pr(F > c \mid H_0) = 0.05$. $\diamond$