1. (a) μ_1 is the mean for treatment 1. Because **R** uses the design matrix

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix},$$

the mean for treatment 1 is estimated by **R**'s estimate of the intercept, i.e., 351.

(b) μ_2 is the mean for treatment 2. This should be clear because our model for the data was stated as $y_{ij} = \mu_i + \epsilon_{ij}$, where ϵ_{ij} terms iid $N(0, \sigma^2)$. In **R**, the mean for treatment 2 is estimated by $\widehat{\text{Intercept}} + \widehat{Dose2}$ (see rows 3 and 4 of design matrix). Thus, $\hat{\mu}_2 = 351 - 10 = 341$.

(c) $\hat{\mu}_2 = \frac{y_{21}+y_{22}}{2} = \bar{y}_{2.}$ and $Var(\hat{\mu}_2) = \frac{\sigma^2}{2}$. Also, $\hat{\sigma}^2 = \frac{SSE}{n-r} = \frac{432.5}{10-5} = \frac{432.5}{5}$. Thus,

$$SE(\hat{\mu}_2) = \sqrt{\widehat{Var}(\hat{\mu}_2)} = \sqrt{\hat{\sigma}^2/2} = \sqrt{\frac{432.5}{(2)(5)}}.$$

I was hoping that many of you would recognize that $SE(\hat{\mu}_1) = \cdots = SE(\hat{\mu}_5)$ due to the balanced design. Thus, $SE(\hat{\mu}_2) = SE(\hat{\mu}_1) = SE(\widehat{\text{Intercept}}) = 6.576$. This is the same as $\sqrt{\frac{432.5}{(2)(5)}}$ derived above. Some of you computed $SE(\hat{\mu}_2)$ by assuming

$$Var(\widehat{\text{Intercept}} + \widehat{\text{Dose2}}) = Var(\widehat{\text{Intercept}}) + Var(\widehat{\text{Dose2}}).$$

The problem is that $\widehat{\text{Intercept}}$ and $\widehat{\text{Dose2}}$ ($\bar{y}_{1.}$ and $\bar{y}_{2.} - \bar{y}_{1.}$, respectively) are not independent.

(d) $\widehat{\text{Dose2}}$ is an estimate of $\mu_2 - \mu_1$. Thus, a t-statistic for testing $H_0 : \mu_1 = \mu_2$ is given in the **R** output

$$\text{t} = -1.075, \ \text{p-value} = 0.3314$$

This indicates that there is no significant evidence of a difference between μ_1 and μ_2. The distribution of this t-statistic is noncentral t with $n - r = 10 - 5 = 5$ *df* and $ncp = \frac{\mu_1 - \mu_2}{\sqrt{\sigma^2\left(\frac{1}{2}+\frac{1}{2}\right)}} = \frac{\mu_1-\mu_2}{\sigma}$. Many of you failed to state the distribution of the test statistic or failed to provide the *df* or *ncp*. I tried to clue you in on this by my statement of "(be very precise)" when I asked for the distribution of the test statistic.

(e) The t-statistic for testing $H_0 : \mu_3 = \mu_4$ is

$$t = \frac{\bar{y}_{3.} - \bar{y}_{4.}}{\sqrt{MSE\left(\frac{1}{2} + \frac{1}{2}\right)}} = \frac{\bar{y}_{3.} - \bar{y}_{4.}}{\sqrt{MSE}}.$$

This can be computed from the **R** output as

$$t = \frac{-6 - (-17)}{\sqrt{432.5/5}} = \frac{11}{\sqrt{432.5/5}}.$$

Thus, the F-statistic is

$$t^2 = \frac{(11)^2}{432.5/5}.$$

You might instead have noticed from the **R** output that the SE of a difference in treatment means (which is the same for all treatment pairs due to the balanced design) is 9.301. This leads to the same F-statistic as that derived above $\left(\frac{11}{9.301}\right)^2$. Some of you gave the general formula for an F-statistic. This problem asked you to "use the **R** code and partial output provided below to answer the following questions." Thus, the point was to see if you understood enough about the relationships of various quantities to come up with a numerical answer.

(f) We need to conduct a test for "lack of fit." This can be done by comparing full and reduced models

$$\begin{aligned} F &= \frac{(SSE_{reduced} - SSE_{full})/(df_{reduced} - df_{full})}{SSE_{full}/df_{full}} \\ &= \frac{(1038.5 - 432.5)/((10-2)-(10-5))}{432.5/(10-5)} \\ &= \frac{(1038.5 - 432.5)/3}{432.5/5}. \end{aligned}$$

This is an F-statistic with *df* 3 and 5. The p-value can be obtained from the last ANOVA table in the output, which shows that p=0.1907591. Thus, there is no significant evidence of lack of linear fit.

(g) The key to this problem is to notice that the Doses are not equally spaced. If the means fall along a line, then there exist β_0 and β_1 such that

$$\begin{array}{l} \mu_1 = \beta_0 + \beta_1(0) \\ \mu_2 = \beta_0 + \beta_1(2) \\ \mu_3 = \beta_0 + \beta_1(4) \\ \mu_4 = \beta_0 + \beta_1(8) \\ \mu_5 = \beta_0 + \beta_1(16) \end{array} \iff \begin{array}{l} \mu_2 - \mu_1 = 2\beta_1 \\ \mu_3 - \mu_2 = 2\beta_1 \\ \mu_4 - \mu_3 = 4\beta_1 \\ \mu_5 - \mu_4 = 8\beta_1 \end{array} \iff \begin{array}{r} -\mu_1 + 2\mu_2 - \mu_3 = 0 \\ 2\mu_2 - 3\mu_3 + \mu_4 = 0 \\ 2\mu_3 - 3\mu_4 + \mu_5 = 0 \end{array}$$

implying that

$$\boldsymbol{C} = \begin{bmatrix} -1 & 2 & -1 & 0 & 0 \\ 0 & 2 & -3 & 1 & 0 \\ 0 & 0 & 2 & -3 & 1 \end{bmatrix} \text{ and } \boldsymbol{d} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

(h) The ANOVA table produced by the **R** command *anova(o3)* is as follows:

Df	Sum of Sq	MS	F
1	5899.6	5899.6	5899.6/(432.5/5)
3	1038.5-432.5	202	202/(432.5/5)
5	432.5	432.5/5	

These entries all follow from the sums of squares $\boldsymbol{y}'(\boldsymbol{P}_2-\boldsymbol{P}_1)\boldsymbol{y}, \boldsymbol{y}'(\boldsymbol{P}_3-\boldsymbol{P}_2)\boldsymbol{y}, \boldsymbol{y}'(\boldsymbol{P}_4-\boldsymbol{P}_3)\boldsymbol{y}$ where

$$\boldsymbol{X}_1 = \mathbf{1}_{10\times 1},\ \boldsymbol{X}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 1 & 2 \\ 1 & 4 \\ 1 & 4 \\ 1 & 8 \\ 1 & 8 \\ 1 & 16 \\ 1 & 16 \end{bmatrix},\ \boldsymbol{X}_3 = \boldsymbol{I}_{5\times 5} \otimes \mathbf{1}_{2\times 1} \text{ and } \boldsymbol{X}_4 = \boldsymbol{I}_{10\times 10}.$$

The *df* are differences in ranks of these matrices, i.e., 2-1, 5-2 and 10-5. Recall that the SS can be seen as reductions in SSE when projecting onto a larger column space compared to projecting onto a smaller column space. For example, $\boldsymbol{y}'(\boldsymbol{P}_3 - \boldsymbol{P}_2)\boldsymbol{y} = \boldsymbol{y}'((\boldsymbol{I} - \boldsymbol{P}_2) - (\boldsymbol{I} - \boldsymbol{P}_3))\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_2)\boldsymbol{y} - \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_3)\boldsymbol{y} = 1038.5 - 432.5 = 606$ is reduction in SSE when projecting on $\boldsymbol{X}_3$ instead of $\boldsymbol{X}_2$.

2. Prove that $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{X}\boldsymbol{B}^{-1})$:

$$\begin{aligned}
\boldsymbol{a} \in \mathcal{C}(\boldsymbol{X}) &\iff \underset{n\times 1}{\boldsymbol{a}} = \boldsymbol{X}\boldsymbol{b} && \text{for some } \underset{p\times 1}{\boldsymbol{b}} \\
&\iff \boldsymbol{a} = \boldsymbol{X} \underset{p\times p}{\boldsymbol{I}} \boldsymbol{b} && \text{for some } \boldsymbol{b}\text{x} \\
&\iff \boldsymbol{a} = \boldsymbol{X}\boldsymbol{B}^{-1} \underbrace{\boldsymbol{B}\boldsymbol{b}}_{p\times 1} && \text{treat as } \boldsymbol{X}\boldsymbol{B}^{-1} \text{ product a } p\times 1 \text{ vector} \\
&\implies \boldsymbol{a} \in \mathcal{C}(\boldsymbol{X}\boldsymbol{B}^{-1})
\end{aligned}$$

So $\mathcal{C}(\boldsymbol{X}) \subseteq \mathcal{C}(\boldsymbol{X}\boldsymbol{B}^{-1})$.
Then similarly,

$$\begin{aligned}
\boldsymbol{g} \in \mathcal{C}(\boldsymbol{X}\boldsymbol{B}^{-1}) &\iff \boldsymbol{g} = \boldsymbol{X}\boldsymbol{B}^{-1}\boldsymbol{h} && \text{for some } p\times 1 \text{ vector } \boldsymbol{h} \\
&\iff \boldsymbol{g} = \boldsymbol{X} \underbrace{\boldsymbol{B}^{-1}\boldsymbol{h}}_{p\times 1} && \text{treat as } \boldsymbol{X} \text{ product a } p\times 1 \text{ vector} \\
&\implies \boldsymbol{g} \in \mathcal{C}(\boldsymbol{X})
\end{aligned}$$

So $\mathcal{C}(\boldsymbol{X}\boldsymbol{B}^{-1}) \subseteq \mathcal{C}(\boldsymbol{X})$.
According to the results above, $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{X}\boldsymbol{B}^{-1})$.

3. Prove that $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{W}}$, i.e. $\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{W}} = \mathbf{0}$:

$$\text{Key: } \mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{W}) \implies \begin{cases} \boldsymbol{X} = \boldsymbol{W}\boldsymbol{A} \text{ for some } \boldsymbol{A} \\ \boldsymbol{W} = \boldsymbol{X}\boldsymbol{B} \text{ for some } \boldsymbol{B} \end{cases}$$

From homework 1 problem 7 (a), we also know that

$$\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{W}} = \mathbf{0} \iff (\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{W}})'(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{W}}) = \mathbf{0}$$

So it is equivalent to prove $(\boldsymbol{P_X}-\boldsymbol{P_W})'(\boldsymbol{P_X}-\boldsymbol{P_W})=\boldsymbol{0}$.

$$\begin{aligned}
(\boldsymbol{P_X}-\boldsymbol{P_W})'(\boldsymbol{P_X}-\boldsymbol{P_W}) &= (\boldsymbol{P_X'}-\boldsymbol{P_W'})(\boldsymbol{P_X}-\boldsymbol{P_W}) \text{ by the property of transpose operation}\\
&= (\boldsymbol{P_X}-\boldsymbol{P_W})(\boldsymbol{P_X}-\boldsymbol{P_W})\\
&= \boldsymbol{P_X}\boldsymbol{P_X}-\boldsymbol{P_X}\boldsymbol{P_W}-\boldsymbol{P_W}\boldsymbol{P_X}+\boldsymbol{P_W}\boldsymbol{P_W}\\
&= \boldsymbol{P_X}-\boldsymbol{P_X}\boldsymbol{W}(\boldsymbol{W'W})^-\boldsymbol{W'}-\boldsymbol{P_W}\boldsymbol{X}(\boldsymbol{X'X})^-\boldsymbol{X'}+\boldsymbol{P_W}\\
&= \boldsymbol{P_X}-\underbrace{\boldsymbol{P_X}\boldsymbol{X}}\boldsymbol{B}(\boldsymbol{W'W})^-\boldsymbol{W'}-\underbrace{\boldsymbol{P_W}\boldsymbol{W}}\boldsymbol{A}(\boldsymbol{X'X})^-\boldsymbol{X'}+\boldsymbol{P_W}\\
&= \boldsymbol{P_X}-\boldsymbol{X}\boldsymbol{B}(\boldsymbol{W'W})^-\boldsymbol{W'}-\boldsymbol{W}\boldsymbol{A}(\boldsymbol{X'X})^-\boldsymbol{X'}+\boldsymbol{P_W}\\
&= \boldsymbol{P_X}-\boldsymbol{P_W}-\boldsymbol{P_X}+\boldsymbol{P_W}\\
&= \boldsymbol{0}
\end{aligned}$$

Therefore the equivalent statement $\boldsymbol{P_X}=\boldsymbol{P_W}$ holds.

4. (a) The corresponding design matrix is

$$\boldsymbol{X}=\begin{pmatrix}1&1&0&1&0\\1&1&0&1&0\\1&1&0&0&1\\1&1&0&0&1\\1&0&1&1&0\\1&0&1&1&0\\1&0&1&0&1\\1&0&1&0&1\end{pmatrix}.$$

(b) Recall that a scalar $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable if there exists a vector $\boldsymbol{a}'$ such that

$$\boldsymbol{c}'\boldsymbol{\beta}=\boldsymbol{a}'\,\mathrm{E}(\boldsymbol{y}),$$

since $\mathrm{E}(\boldsymbol{y})=\boldsymbol{X}\boldsymbol{\beta}$. Hence, $\tau_1-\tau_2$ is estimable if it can be written as a linear combination of the expected value of $\boldsymbol{y}$. Put

$$\boldsymbol{a}'=(1,1,-1,-1,1,1,-1,-1)/4.$$

Then

$$\begin{aligned}
\boldsymbol{a}'\,\mathrm{E}(\boldsymbol{y}) &= \frac{1}{4}\sum_{i=1}^{2}\sum_{k=1}^{2}\big[\mathrm{E}(y_{i1k})-\mathrm{E}(y_{i2k})\big]\\
&= \frac{1}{4}\sum_{i=1}^{2}\sum_{k=1}^{2}\big[\mathrm{E}(\mu+\lambda_i+\tau_1+\varepsilon_{i1k})-\mathrm{E}(\mu+\lambda_i+\tau_2+\varepsilon_{i2k})\big]\\
&= \frac{1}{4}\sum_{i=1}^{2}\sum_{k=1}^{2}\big[(\mu+\lambda_i+\tau_1+0)-(\mu+\lambda_i+\tau_2+0)\big]\\
&= \frac{1}{4}(2\cdot 2)\big[\tau_1-\tau_2\big]\\
&= \tau_1-\tau_2.
\end{aligned}$$

Alternatively, and more simply, $\mathrm{E}(y_{111})-\mathrm{E}(y_{121})=\tau_1-\tau_2$, so $\tau_1-\tau_2$ is estimable.

(c) Notice that rank($\boldsymbol{X}$) = 3, so that $\boldsymbol{X}^*$ needs to be 8×3 to have full column rank and the same column space as $\boldsymbol{X}$. One possible choice is

$$\boldsymbol{X}^* = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}.$$

(d) Let $\boldsymbol{\beta}^* = (\beta_1^*, \beta_2^*, \beta_3^*)'$. Using $\boldsymbol{X}^*$ as given in part (c), equating expected values gives

$$\begin{aligned} \beta_1^* + \beta_2^* + \beta_3^* &= \mu + \lambda_1 + \tau_1, \\ \beta_1^* + \beta_2^* - \beta_3^* &= \mu + \lambda_1 + \tau_2, \\ \beta_1^* - \beta_2^* + \beta_3^* &= \mu + \lambda_2 + \tau_1, \\ \beta_1^* - \beta_2^* - \beta_3^* &= \mu + \lambda_2 + \tau_2. \end{aligned}$$

Then

$$\begin{aligned} 2\beta_1^* &= 2\mu + \lambda_1 + \lambda_2 + \tau_1 + \tau_2, \\ 2\beta_2^* &= \lambda_1 - \lambda_2, \\ 2\beta_3^* &= \tau_1 - \tau_2, \end{aligned}$$

which implies

$$\begin{aligned} \beta_1^* &= \mu + (\lambda_1 + \lambda_2)/2 + (\tau_1 + \tau_2)/2, \\ \beta_2^* &= (\lambda_1 - \lambda_2)/2, \\ \beta_3^* &= (\tau_1 - \tau_2)/2. \end{aligned}$$

(e) Since $\tau_1 - \tau_2 = 2\beta_3^* = \left(0,\ 0,\ 2\right)\boldsymbol{\beta}^*$,

$$
\begin{aligned}
\left(\widehat{\tau_1 - \tau_2}\right)_{\text{OLS}} &= \left(0,\ 0,\ 2\right)\hat{\boldsymbol{\beta}}^*_{\text{OLS}} \\
&= \left(0,\ 0,\ 2\right)(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{y} \\
&= \left(0,\ 0,\ 2\right)\begin{pmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{pmatrix}^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{y} \\
&= \left(0,\ 0,\ 2\right)\begin{pmatrix} 1/8 & 0 & 0 \\ 0 & 1/8 & 0 \\ 0 & 0 & 1/8 \end{pmatrix}\boldsymbol{X}^{*\prime}\boldsymbol{y} \\
&= \left(0,\ 0,\ 1/4\right)\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \end{pmatrix}\boldsymbol{y} \\
&= \frac{1}{4}\begin{pmatrix} 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \end{pmatrix}\begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{pmatrix} \\
&= \frac{1}{4}\sum_{i=1}^{2}\sum_{k=1}^{2}\left[y_{i1k} - y_{i2k}\right] \\
&= \bar{y}_{\cdot 1\cdot} - \bar{y}_{\cdot 2\cdot}.
\end{aligned}
$$