1. $[\mathbf{p}_1'\boldsymbol{y}, \mathbf{p}_2'\boldsymbol{y}, ..., \mathbf{p}_n'\boldsymbol{y}]' = \boldsymbol{P}'\boldsymbol{y}$ where $\boldsymbol{P} = [\mathbf{p}_1, ..., \mathbf{p}_n]$. Because linear transformations of multivariate normal vectors are multivariate normal, we know

$$\boldsymbol{P}'\boldsymbol{y} \sim N(E(\boldsymbol{P}'\boldsymbol{y}), Var(\boldsymbol{P}'\boldsymbol{y}))$$

where $E(\boldsymbol{P}'\boldsymbol{y}) = \boldsymbol{P}'E(\boldsymbol{y}) = \boldsymbol{P}'\mathbf{0} = \mathbf{0}$ and $Var(\boldsymbol{P}'\boldsymbol{y}) = \boldsymbol{P}'Var(\boldsymbol{y})\boldsymbol{P} = \boldsymbol{P}'\boldsymbol{\Sigma}\boldsymbol{P}$. By the spectral decomposition theorem, we know $\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$ where $\boldsymbol{\Lambda} = Diag(\lambda_1, ..., \lambda_n)$.Thus, $Var(\boldsymbol{P}'\boldsymbol{y}) = \boldsymbol{P}'\boldsymbol{\Sigma}\boldsymbol{P} = \boldsymbol{P}'\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}\boldsymbol{\Lambda}\boldsymbol{I} = \boldsymbol{\Lambda}$. So, $\boldsymbol{P}'\boldsymbol{y} \sim N(\mathbf{0}, \boldsymbol{\Lambda})$.

2. (a) A linear-mixed effects model for the overall quality score is

$$y_{ijk} = \mu + \alpha_i + u_{ij} + \epsilon_{ijk},$$

where

- α_i is the fixed effect corresponding to temperature level $i = 1, 2, 3$,
- u_{ij} is the random effect corresponding to cooler $j = 1, 2, 3, 4$ at temperature level i,
- ϵ_{ijk} is the random error for beef cut $k = 1, 2$ in cooler j at temperature level i, and
- $u_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_u^2)$ independent of $\epsilon_{ijk} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$.

In matrix form, this model is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon},$$

where

- $\boldsymbol{y} = (y_{111}, y_{112}, y_{121}, \ldots, y_{142}, y_{211}, \ldots, y_{342})'$,
- $\boldsymbol{X} = (\mathbf{1}_{24\times1}, \boldsymbol{I}_{3\times3} \otimes \mathbf{1}_{8\times1})$,
- $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \alpha_3)'$,
- $\boldsymbol{Z} = (\boldsymbol{I}_{12\times12} \otimes \mathbf{1}_{2\times1})$,
- $\boldsymbol{u} = (u_{11}, u_{12}, \ldots, u_{34})'$,
- $\boldsymbol{\epsilon} = (\epsilon_{111}, \epsilon_{112}, \epsilon_{121}, \ldots, \epsilon_{142}, \epsilon_{211}, \ldots, \epsilon_{342})'$, and
- $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0}_{12\times1} \\ \mathbf{0}_{24\times1} \end{pmatrix}, \begin{pmatrix} \sigma_u^2\boldsymbol{I}_{12\times12} & \mathbf{0}_{12\times24} \\ \mathbf{0}_{24\times12} & \sigma_\epsilon^2\boldsymbol{I}_{24\times24} \end{pmatrix} \right)$.

(b) ANOVA table:

Source	DF	Sums of Squares	Mean Squares	Expected Mean Squares
temperature	3–1=2	$\sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{k=1}^{2}(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot\cdot\cdot})^2$	$\frac{8}{2}\sum_{i=1}^{3}(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot\cdot\cdot})^2$	$\sigma_\epsilon^2+2\sigma_u^2+4\sum_{i=1}^{3}(\alpha_i-\bar{\alpha}_\cdot)^2$
cooler(temp)	(4–1)(3)=9	$\sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{k=1}^{2}(\bar{y}_{ij\cdot}-\bar{y}_{i\cdot\cdot})^2$	$\frac{2}{9}\sum_{i=1}^{3}\sum_{j=1}^{4}(\bar{y}_{ij\cdot}-\bar{y}_{i\cdot\cdot})^2$	$\sigma_\epsilon^2+2\sigma_u^2$
cut(cooler,temp)	(2–1)(3)(4)=12	$\sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{k=1}^{2}(y_{ijk}-\bar{y}_{ij\cdot})^2$	$\frac{1}{12}\sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{k=1}^{2}(y_{ijk}-\bar{y}_{ij\cdot})^2$	σ_ϵ^2
c. total	24–1=23	$\sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{k=1}^{2}(y_{ijk}-\bar{y}_{\cdot\cdot\cdot})^2$		

(c) A test of $H_0 : \alpha_1 - \alpha_2 = 0$ can be based on

$$t = \frac{\bar{y}_{1\cdot\cdot} - \bar{y}_{2\cdot\cdot} - 0}{\sqrt{\frac{2MS_{\text{cooler(temp)}}}{4\cdot 2}}} = \frac{\bar{y}_{1\cdot\cdot} - \bar{y}_{2\cdot\cdot}}{\sqrt{\frac{1}{4}\left(\frac{2}{9}\sum_{i=1}^{3}\sum_{j=1}^{4}(\bar{y}_{ij\cdot}-\bar{y}_{i\cdot\cdot})^2\right)}} = \frac{\bar{y}_{1\cdot\cdot} - \bar{y}_{2\cdot\cdot}}{\sqrt{\frac{1}{18}\sum_{i=1}^{3}\sum_{j=1}^{4}(\bar{y}_{ij\cdot}-\bar{y}_{i\cdot\cdot})^2}}.$$

The numerator should be obvious, but why use $\frac{2MS_{\text{cooler(temp)}}}{4\cdot 2}$ in the denominator? Notice that since the u_{ij} and e_{ijk} are all independent,

$$\begin{aligned}
\text{Var}(\bar{y}_{1\cdot\cdot} - \bar{y}_{2\cdot\cdot}) &= \text{Var}(\bar{y}_{1\cdot\cdot}) + \text{Var}(\bar{y}_{2\cdot\cdot}) - 2\,\text{Cov}(\bar{y}_{1\cdot\cdot}, \bar{y}_{2\cdot\cdot}) \\
&= \text{Var}\left(\mu + \alpha_1 + \bar{u}_{1\cdot} + \bar{\epsilon}_{1\cdot\cdot}\right) + \text{Var}\left(\mu + \alpha_2 + \bar{u}_{2\cdot} + \bar{\epsilon}_{2\cdot\cdot}\right) \\
&\qquad - 2\,\text{Cov}\left(\mu + \alpha_1 + \bar{u}_{1\cdot} + \bar{\epsilon}_{1\cdot\cdot}, \mu + \alpha_2 + \bar{u}_{2\cdot} + \bar{\epsilon}_{2\cdot\cdot}\right) \\
&= \text{Var}\left(\bar{u}_{1\cdot} + \bar{\epsilon}_{1\cdot\cdot}\right) + \text{Var}\left(\bar{u}_{2\cdot} + \bar{\epsilon}_{2\cdot\cdot}\right) - 2\,\text{Cov}\left(\bar{u}_{1\cdot} + \bar{\epsilon}_{1\cdot\cdot}, \bar{u}_{2\cdot} + \bar{\epsilon}_{2\cdot\cdot}\right) \\
&= \text{Var}\left(\bar{u}_{1\cdot}\right) + \text{Var}\left(\bar{\epsilon}_{1\cdot\cdot}\right) + \text{Var}\left(\bar{u}_{2\cdot}\right) + \text{Var}\left(\bar{\epsilon}_{2\cdot\cdot}\right) \\
&= \frac{\sigma_u^2}{4} + \frac{\sigma_\epsilon^2}{2\cdot 4} + \frac{\sigma_u^2}{4} + \frac{\sigma_\epsilon^2}{2\cdot 4} \\
&= \frac{2\left(\sigma_\epsilon^2 + 2\sigma_u^2\right)}{4\cdot 2} \\
&= \frac{2EMS_{\text{cooler(temp)}}}{4\cdot 2}.
\end{aligned}$$

(d) The degrees of freedom are 9, since the denominator is based on $MS_{\text{cooler(temp)}}$.

(e) The noncentrality parameter is

$$\frac{\alpha_1 - \alpha_2 - 0}{\sqrt{\frac{2(\sigma_\epsilon^2 + 2\sigma_u^2)}{4\cdot 2}}} = \frac{2(\alpha_1 - \alpha_2)}{\sqrt{\sigma_\epsilon^2 + 2\sigma_u^2}}.$$

3. (a) The covariance between the heights of two plants (i.e., genotypes $k = 1, 2$) on the same table (i.e., watering level j and greenhouse i) is

$$\begin{aligned}
\text{Cov}(y_{ij1}, y_{ij2}) &= \text{Cov}(\mu + g_i + \omega_j + t_{ij} + \gamma_1 + \phi_{j1} + e_{ij1}, \mu + g_i + \omega_j + t_{ij} + \gamma_2 + \phi_{j2} + e_{ij2}) \\
&= \text{Cov}(g_i + t_{ij} + e_{ij1}, g_i + t_{ij} + e_{ij2}) \qquad \text{dropping fixed effects} \\
&= \text{Cov}(g_i, g_i) + \text{Cov}(t_{ij}, t_{ij}) \qquad \text{since } g_i, t_{ij}, e_{ijk} \text{ are all independent} \\
&= \sigma_g^2 + \sigma_t^2.
\end{aligned}$$

The variance of any single observation is

$$\begin{aligned}
\text{Var}(y_{ijk}) &= \text{Var}(\mu + g_i + \omega_j + t_{ij} + \gamma_k + \phi_{jk} + e_{ijk}) \\
&= \text{Cov}(g_i + t_{ij} + e_{ijk}, g_i + t_{ij} + e_{ijk}) \qquad \text{dropping fixed effects} \\
&= \text{Cov}(g_i, g_i) + \text{Cov}(t_{ij}, t_{ij}) + \text{Cov}(e_{ijk}, e_{ijk}) \qquad \text{since } g_i, t_{ij}, e_{ijk} \text{ are all independent} \\
&= \sigma_g^2 + \sigma_t^2 + \sigma_e^2.
\end{aligned}$$

Hence, the correlation is

$$\begin{aligned}
\text{Corr}(y_{ij1}, y_{ij2}) &= \frac{\text{Cov}(y_{ij1}, y_{ij2})}{\sqrt{\text{Var}(y_{ij1})\,\text{Var}(y_{ij2})}} \\
&= \frac{\sigma_g^2 + \sigma_t^2}{\sigma_g^2 + \sigma_t^2 + \sigma_e^2}.
\end{aligned}$$

(b) If there are no watering level main effects, the fixed effects will be the same for each watering level j when averaged across the other factors (i.e., averaged over i and k). Written in terms of the model parameters, $\mu + \omega_j + \bar{\gamma}_{\cdot} + \bar{\phi}_{j\cdot}$ would be equal for all j. This happens if and only if $\omega_j + \bar{\phi}_{j\cdot}$ is equal for all j, so the null hypothesis of no watering level main effects is

$$H_0 : \omega_1 + \bar{\phi}_{1\cdot} = \omega_2 + \bar{\phi}_{2\cdot} = \omega_3 + \bar{\phi}_{3\cdot}$$

Comments: Note that $H_0 : \omega_1 = \omega_2 = \omega_3$ is *not* the null hypothesis of no watering level main effects. Even if $\omega_1 = \omega_2 = \omega_3$, there could still be main effects from the interaction terms.

(c) Let

- $\boldsymbol{\beta} = (\mu, \omega_1, \omega_2, \omega_3, \gamma_1, \gamma_2, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \phi_{31}, \phi_{32})'$,
- $\boldsymbol{X} = (\mathbf{1}_{24\times1}, \mathbf{1}_{4\times1} \otimes \boldsymbol{I}_{3\times3} \otimes \mathbf{1}_{2\times1}, \mathbf{1}_{12\times1} \otimes \boldsymbol{I}_{2\times2}, \mathbf{1}_{4\times1} \otimes \boldsymbol{I}_{6\times6})$,
- $\boldsymbol{u} = (g_1, g_2, g_3, g_4, t_{11}, t_{12}, t_{13}, t_{21}, \ldots, t_{43})'$,
- $\boldsymbol{Z} = (\boldsymbol{I}_{4\times4} \otimes \mathbf{1}_{6\times1}, \boldsymbol{I}_{12\times12} \otimes \mathbf{1}_{2\times1})$.

4. For this problem, let μ_{ijk} denote the mean for filler type $i = 1, 2$, surface treatment $j = 1, 2$, and filler proportion $k = 1, 2, 3$ (corresponding to 25%, 50%, and 75%, respectively).

The `R` code below fits the cell means model to these data:

```
> # Read in data and set covariates as factors
> dat <- read.table("https://dnett.github.io/S510/FabricLoss.txt",header = T,
+                   col.names=c("S","F","p","y"),
+                   colClasses = c("factor","factor","factor","numeric"))

> # Fit cell means model.
> fit <- lm(y ~ F:S:p + 0, data = dat)
```

From the output below, notice that `R` is using the parameterization

$$\boldsymbol{\beta} = [\mu_{111}, \mu_{211}, \mu_{121}, \mu_{221}, \mu_{112}, \mu_{212}, \mu_{122}, \mu_{222}, \mu_{113}, \mu_{213}, \mu_{123}, \mu_{223}]'. \tag{1}$$

```
> # Look at the vector of estimates.
> coef(fit)
F1:S1:p25 F2:S1:p25 F1:S2:p25 F2:S2:p25 F1:S1:p50 F2:S1:p50
    201.0     213.0     164.0     148.5     237.0     233.5
F1:S2:p50 F2:S2:p50 F1:S1:p75 F2:S1:p75 F1:S2:p75 F2:S2:p75
    187.5     113.5     267.0     234.5     232.0     143.5
```

Depending how you called `lm()`, you may not have the same parameterization of $\boldsymbol{\beta}$ as in (1), and hence may require different $\boldsymbol{C}$ matrices in the following problems.

(a) From the output below,

$$\hat{\mu}_{212} = 233.50 \quad \text{and} \quad \text{se}(\hat{\mu}_{212}) = 11.59.$$

```
> # 4a: mean and se for F2, S1, 50% filler.
> summary(fit)

Call:
lm(formula = y ~ F:S:p + 0, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-26.000  -9.125   0.000   9.125  26.000

Coefficients:
          Estimate Std. Error t value Pr(>|t|)
F1:S1:p25   201.00      11.59  17.340 7.33e-10 ***
F2:S1:p25   213.00      11.59  18.375 3.74e-10 ***
F1:S2:p25   164.00      11.59  14.148 7.57e-09 ***
F2:S2:p25   148.50      11.59  12.811 2.33e-08 ***
F1:S1:p50   237.00      11.59  20.445 1.08e-10 ***
F2:S1:p50   233.50      11.59  20.143 1.28e-10 ***
F1:S2:p50   187.50      11.59  16.175 1.64e-09 ***
F2:S2:p50   113.50      11.59   9.791 4.50e-07 ***
F1:S1:p75   267.00      11.59  23.033 2.67e-11 ***
F2:S1:p75   234.50      11.59  20.229 1.22e-10 ***
F1:S2:p75   232.00      11.59  20.014 1.38e-10 ***
F2:S2:p75   143.50      11.59  12.379 3.42e-08 ***
```

(b) From the code and output below (which uses Dr. Nettleton's `estimate()` function),

$$\widehat{\bar{\mu}}_{1\cdot\cdot} = 214.75 \quad \text{and} \quad \widehat{\bar{\mu}}_{2\cdot\cdot} = 181.08.$$

```
> # 4b: LSMEANS for F1 and F2.
> C.b <- c(rep(c(1,0), 6),    # F1
+          rep(c(0,1), 6))    # F2
> C.b <- matrix(C.b / 6, nrow = 2, byrow = TRUE)
> estimate(fit, C.b)
            c1        c2        c3        c4        c5        c6        c7        c8
[1,] 0.1666667 0.0000000 0.1666667 0.0000000 0.1666667 0.0000000 0.1666667 0.0000000
[2,] 0.0000000 0.1666667 0.0000000 0.1666667 0.0000000 0.1666667 0.0000000 0.1666667
            c9       c10       c11       c12 estimate       se 95% Conf.   limits
[1,] 0.1666667 0.0000000 0.1666667 0.0000000 214.7500 4.732424  204.4389 225.0611
[2,] 0.0000000 0.1666667 0.0000000 0.1666667 181.0833 4.732424  170.7723 191.3944
```

(c) From the code and output below (which uses Dr. Nettleton's `estimate()` function),

$$\widehat{\bar{\mu}}_{\bullet 21} = 156.25.$$

```
> # 4c-d. LSMEAN and se for S2&p25.
> C.c <- c(0,0, 1, 1, rep(0, 8))
> C.c <- matrix(C.c / 2, nrow = 1)
> estimate(fit, C.c)
     c1 c2  c3  c4 c5 c6 c7 c8 c9 c10 c11 c12 estimate       se 95% Conf.   limits
[1,]  0  0 0.5 0.5  0  0  0  0  0   0   0   0   156.25 8.196798  138.3907 174.1093
```

(d) From the output in part (c),

$$\text{se}(\widehat{\bar{\mu}}_{\bullet 12}) = 8.197.$$

(e) The test for filler type main effects is given by

$$H_0 : \bar{\mu}_{1\bullet\bullet} - \bar{\mu}_{2\bullet\bullet} = 0.$$

The code below (which uses Dr. Nettleton's `test()` function) gives $F = 25.30$ and $p = 0.0003$. These data provide strong evidence that there are filler type main effects. That is, averaging over surface treatment and proportion of filler, there is strong evidence that filler type F1 differs from F2 at preventing fabric loss in abrasive tests.

```
> # 4e. Filler type main effects.
> C.e <- matrix(rep(c(1, -1), 6), nrow = 1)
> test(fit, C.e) # Note d = 0 by default.
$Fstat
[1] 25.30481
$pvalue
[1] 0.0002939847
```

Comments: we are specifically interested in main effects here, not any kind of effect. Hence, comparing a reduced cell-means model with only the factors surface treatment and filler proportion to a full cell-means model with all three factors does not test for filler type main effects.

(f) The test for three-way interactions is given by

$$H_0 : \begin{bmatrix} \mu_{111} - \mu_{121} - \mu_{211} + \mu_{221} - \mu_{112} + \mu_{122} + \mu_{212} - \mu_{222} \\ \mu_{111} - \mu_{121} - \mu_{211} + \mu_{221} - \mu_{113} + \mu_{123} + \mu_{213} - \mu_{223} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The code below (which uses Dr. Nettleton's `test()` function) gives $F = 0.89$ and $p = 0.44$. These data do not provide evidence of three-way interactions among filler type, surface treatment, and filler proportion.

```
> # 4f. Three-factor interaction.
> C.f <- c(1, -1, -1, 1, -1, 1, 1, -1, 0, 0, 0, 0,
+          1, -1, -1, 1, 0, 0, 0, 0, -1, 1, 1, -1)
> C.f <- matrix(C.f, nrow = 2, byrow = TRUE)
> test(fit, C.f)
$Fstat
[1] 0.8903876
$pvalue
[1] 0.4359589
```

Comments: a large p-value (i.e., $p > \alpha$) suggests that it *possible* that there are no three-way interactions, but it does not say there are no three-way interactions. Why? There could be three-way interactions, but they are too small (relative to the variability in the study) to detect.

(g) The test for two-way interactions between filler type and filler proportion is given by

$$H_0 : \begin{bmatrix} \bar{\mu}_{1.1} - \bar{\mu}_{1.2} - \bar{\mu}_{2.1} + \bar{\mu}_{2.2} \\ \bar{\mu}_{1.1} - \bar{\mu}_{1.3} - \bar{\mu}_{2.1} + \bar{\mu}_{2.3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The code below (which uses Dr. Nettleton's `test()` function) gives $F = 6.57$ and $p = 0.012$. These data provide evidence that there are two-way interactions between filler type and filler proportion.

```
> # 4g. Two-factor interaction between filler type and proportion.
> C.g <- c(1, -1, 1, -1, -1, 1, -1, 1, 0, 0, 0, 0,
+          1, -1, 1, -1, 0, 0, 0, 0, -1, 1, -1, 1)
> C.g <- matrix(C.g, nrow = 2, byrow = TRUE)
> test(fit, C.g)
$Fstat
[1] 6.565736
$pvalue
[1] 0.01185168
```