AMS256 Homework 4

1. In vector form, our model can be written as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\boldsymbol{y} = (17, 20, 15, 20, 12, 11, 14, 6, 17, 9, 4, 6, 19)^T$, $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_3)^T$, and

$$\underset{13\times 8}{\boldsymbol{X}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix},$$

with rank $r = 6$.

The hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3$ can be re-written as the following two hypotheses: $\beta_1 - \beta_2 = 0$ and $\beta_1 - \beta_3 = 0$ represented as $\boldsymbol{\lambda}_1^T\boldsymbol{\beta} = 0$ and $\boldsymbol{\lambda}_2^T\boldsymbol{\beta} = 0$ where $\boldsymbol{\lambda}_1 = (0, 0, 0, 0, 0, 1, -1, 0)^T$, $\boldsymbol{\lambda}_2 = (0, 0, 0, 0, 0, 1, 0, -1)^T$. This hypothesis is testable if $\boldsymbol{\lambda}_1^T\boldsymbol{\beta}$ and $\boldsymbol{\lambda}_2^T\boldsymbol{\beta}$ are estimable and $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are linearly independent. Note that $\boldsymbol{\lambda}_1^T = \boldsymbol{a}_1^T\boldsymbol{X}$ where $\boldsymbol{a}_1 = (1, 0, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^T$ and $\boldsymbol{\lambda}_2^T = \boldsymbol{a}_2^T\boldsymbol{X}$ where $\boldsymbol{a}_2 = (1, 0, 0, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0)^T$, meaning we get $\boldsymbol{\lambda}_1$ by subtracting row 3 from row 1 in $\boldsymbol{X}$ and we get $\boldsymbol{\lambda}_2$ by subtracting row 4 from row 1 in $\boldsymbol{X}$. Hence, these are estimable functions and the hypothesis is testable.

Now let $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2]$. We write H_0 as $\boldsymbol{\Lambda}^T\boldsymbol{\beta} = \boldsymbol{0}$. If this hypothesis is true, then we have

$$\frac{Q}{\sigma^2} = \frac{(\boldsymbol{\Lambda}^T\hat{\boldsymbol{\beta}} - \boldsymbol{0})^T(\boldsymbol{\Lambda}^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{\Lambda})^{-1}(\boldsymbol{\Lambda}^T\hat{\boldsymbol{\beta}} - \boldsymbol{0})}{\sigma^2} \sim \chi^2_s\,,$$

where $s = \text{rank}(\boldsymbol{\Lambda}) = 2$, $\hat{\boldsymbol{\beta}}$ is any solution to the normal equations and $(\boldsymbol{X}^T\boldsymbol{X})^-$ is any generalized inverse of $(\boldsymbol{X}^T\boldsymbol{X})$. Since Q is independent of $SSE = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$, we also have (if H_0 is true)

$$F = \frac{Q/s}{SSE/(13 - r)} \sim F(s, 13 - r)\,.$$

In this example, $Q = 145.65$, $SSE = 70.51$, and $F = 7.229$. This exceeds the critical value $F_{0.05}(2, 7) = 4.256$, so we reject H_0 at the .05 level and conclude that at least one β_j is not equal to the others.

We can also test this hypothesis in `R`, although we have to account for different parameterization. By default in `R`, the model is written as

$$y_{i,j,k} = \gamma + a_i + b_j + \epsilon_{i,j,k}, \quad i = 1, 2, 3, 4, \quad j = 1, 2, 3\,,$$

where $a_1 = b_1 = 0$, so that

$$
\begin{aligned}
\gamma &= \mu + \alpha_1 + \beta_1 = \mathrm{E}(y_{1,1,1}) \,, \\
a_2 &= (\mu + \alpha_2 + \beta_1) - (\mu + \alpha_1 + \beta_1) \,, \\
&= \alpha_2 - \alpha_1 = \mathrm{E}(y_{2,1,1} - y_{1,1,1}) \,, \\
a_3 &= \alpha_3 - \alpha_1 = \mathrm{E}(y_{3,1,1} - y_{1,1,1}) \,, \\
a_4 &= \alpha_4 - \alpha_1 = \mathrm{E}(y_{4,1,1} - y_{1,1,1}) \,, \\
b_2 &= (\mu + \alpha_1 + \beta_2) - (\mu + \alpha_1 + \beta_1) \,, \\
&= \beta_2 - \beta_1 = \mathrm{E}(y_{1,2,1} - y_{1,1,1}) \,, \\
b_3 &= \beta_3 - \beta_1 = \mathrm{E}(y_{1,3,1} - y_{1,1,1}) \,.
\end{aligned}
$$

With this parameterization, the hypothesis becomes H_0: $b_2 = b_3 = 0$ which can be expressed with $b_2 = 0$ and $b_3 - b_2 = 0$. The following code demonstrates setting up and testing this hypothesis in three ways.

```
y <- c(17, 20, 15, 20, 12, 11, 14, 6, 17, 9, 4, 6, 19)
A <- c(rep(1, 4), rep(2, 3), rep(3, 2), rep(4,4))
B <- c(1,1,2,3, 1,3,3, 1,3, 1,2,2,3)

dat <- data.frame(y=y, A=factor(A), B=factor(B))

mod <- lm(y ~ A + B, data=dat)
model.matrix(mod)
summary(mod)
anova(mod)  # note that this hypothesis is the same as
            # factor B being significant

library(multcomp)
test0 <- glht(mod, linfct=c("B2=0", "B3-B2=0"))
summary(test0, test=Ftest())

(Lam <- matrix(c(0,0,0,0,1,0,
                 0,0,0,0,1,-1),
               nrow=2, byrow=TRUE))
test1 <- glht(mod, linfct=Lam, rhs=c(0,0))
summary(test1, test=Ftest()) # gives the same answer

> General Linear Hypotheses
>
> Linear Hypotheses:
>               Estimate
> B2 == 0         -4.629
> B3 - B2 == 0    9.827
>
> Global Test:
```

```
>    F        DF1 DF2  Pr(>F)
>    7.229    2   7    0.01983
```

2. We can write this model as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\boldsymbol{\beta} = (\alpha_1, \alpha_2)^T$,

$$\underset{3\times 2}{\boldsymbol{X}} = \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{bmatrix}, \ \boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 6 & 0 \\ 0 & 5 \end{bmatrix}, \ \text{and} \ (\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} 1/6 & 0 \\ 0 & 1/5 \end{bmatrix}.$$

We can also write $H_0 : \alpha_1 - \alpha_2 = 0$ which is $\boldsymbol{\lambda}^T\boldsymbol{\beta} = 0$ with $\boldsymbol{\lambda} = (1, -1)^T = \boldsymbol{X}^T\boldsymbol{a}$ and $\boldsymbol{a} = (-1, 1, 0)^T$, so H_0 is testable.

Now,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = \begin{bmatrix} 1/6 & 0 \\ 0 & 1/5 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{6}(y_1 + 2y_2 + y_3) \\ \frac{1}{5}(-y_2 + 2y_3) \end{bmatrix},$$

$$\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1/6 & 0 \\ 0 & 1/5 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{13}{15} & \frac{-1}{15} \\ \frac{1}{6} & \frac{-1}{15} & \frac{29}{30} \end{bmatrix},$$

$$\boldsymbol{I} - \boldsymbol{P} = \begin{bmatrix} \frac{-5}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{-2}{15} & \frac{-1}{15} \\ \frac{1}{6} & \frac{-1}{15} & \frac{-1}{30} \end{bmatrix},$$

$$\boldsymbol{\lambda}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{\lambda} = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1/6 & 0 \\ 0 & 1/5 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{6} + \frac{1}{5} = \frac{11}{30},$$

$$\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} = \frac{1}{6}(y_1 + 2y_2 + y_3) - \frac{1}{5}(-y_2 + 2y_3),$$

$$\begin{aligned} Q &= (\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} - 0)^T(\boldsymbol{\lambda}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{\lambda})^{-1}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} - 0) = \frac{30}{11}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}})^2 \\ &= \frac{1}{330}\left(25y_1^2 + 256y_2^2 + 49y_3^2 + 160y_1y_2 - 170y_1y_3 - 224y_2y_3\right), \end{aligned}$$

and

$$\begin{aligned} SSE = \boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y} &= \frac{1}{30}\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} \begin{bmatrix} -25 & 10 & 5 \\ 10 & -4 & -2 \\ 5 & -2 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \\ &= \frac{1}{30}\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} \begin{bmatrix} -25 & 20 & 10 \\ 0 & -4 & -4 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \\ &= \frac{1}{30}\left(-25y_1^2 - 4y_2^2 - 2y_3^2 + 20y_1y_2 + 10y_1y_3 - 4y_2y_3\right). \end{aligned}$$

Under H_0 :

$$F = \frac{Q/1}{SSE/1} \sim F(1, 3-2),$$

which we can use to test the hypothesis.

3. We can write the model as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\beta} = (\beta_{1,0}, \beta_{1,1}, \beta_{2,0}, \beta_{2,1})^T$ and

$$\underset{(n+m)\times 4}{\boldsymbol{X}} = \begin{bmatrix} \underset{n\times 2}{\boldsymbol{X}_1} & \underset{n\times 2}{\mathbf{0}} \\ \underset{m\times 2}{\mathbf{0}} & \underset{m\times 2}{\boldsymbol{X}_2} \end{bmatrix},$$

where the first columns of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ contain ones and the second columns contain x values. Now

$$\underset{4\times 4}{\boldsymbol{X}^T\boldsymbol{X}} = \begin{bmatrix} \underset{2\times n}{\boldsymbol{X}_1^T} & \underset{2\times m}{\mathbf{0}} \\ \underset{2\times n}{\mathbf{0}} & \underset{2\times m}{\boldsymbol{X}_2^T} \end{bmatrix} \begin{bmatrix} \underset{n\times 2}{\boldsymbol{X}_1} & \underset{n\times 2}{\mathbf{0}} \\ \underset{m\times 2}{\mathbf{0}} & \underset{m\times 2}{\boldsymbol{X}_2} \end{bmatrix} = \begin{bmatrix} \underset{2\times 2}{\boldsymbol{X}_1^T\boldsymbol{X}_1} & \underset{2\times 2}{\mathbf{0}} \\ \underset{2\times 2}{\mathbf{0}} & \underset{2\times 2}{\boldsymbol{X}_2^T\boldsymbol{X}_2} \end{bmatrix},$$

and

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1} & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{X}_2^T\boldsymbol{X}_2)^{-1} \end{bmatrix},$$

so that

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{MLE}} &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \\ &= \begin{bmatrix} (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1} & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{X}_2^T\boldsymbol{X}_2)^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1^T & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_2^T \end{bmatrix} \boldsymbol{y} \\ &= \begin{bmatrix} (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{X}_2^T\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^T \end{bmatrix} \boldsymbol{y} \\ &= \begin{bmatrix} (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{y}_1 \\ (\boldsymbol{X}_2^T\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^T\boldsymbol{y}_2 \end{bmatrix}, \end{aligned}$$

where $\boldsymbol{y}_1$ contains the first n values of $\boldsymbol{y}$ and $\boldsymbol{y}_2$ contains the last m. These are simple linear regressions, so if we denote the first n (x, y) pairs as $(x_{1,i}, y_{1,i})$ for $i = 1, \ldots, n$ and the last m as $(x_{2,j}, y_{2,j})$ for $j = 1, \ldots, m$, then we can write the MLEs as

$$\hat{\beta}_{1,1} = \frac{\sum_{i=1}^n x_{1,i}y_{1,i} - n\bar{x}_1\bar{y}_1}{\sum_{i=1}^n x_{1,i}^2 - n\bar{x}_1}, \ \hat{\beta}_{1,0} = \bar{y}_1 - \hat{\beta}_{1,1}\bar{x}_1,$$

and

$$\hat{\beta}_{2,1} = \frac{\sum_{j=1}^m x_{2,j}y_{2,j} - m\bar{x}_2\bar{y}_2}{\sum_{j=1}^m x_{2,j}^2 - m\bar{x}_2}, \ \hat{\beta}_{2,0} = \bar{y}_2 - \hat{\beta}_{2,1}\bar{x}_2.$$

Also

$$\hat{\sigma}^2_{\text{MLE}} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})/(n+m).$$

As for γ, we know it solves $\beta_{1,0} + \beta_{1,1}\gamma = \beta_{2,0} + \beta_{2,1}\gamma$, so

$$\gamma = \frac{\beta_{1,0} - \beta_{2,0}}{\beta_{2,1} - \beta_{1,1}}.$$

By the invariance property of MLEs, we have

$$\hat{\gamma}_{\text{MLE}} = \frac{\hat{\beta}_{1,0} - \hat{\beta}_{2,0}}{\hat{\beta}_{2,1} - \hat{\beta}_{1,1}}.$$

Now we find the confidence interval of γ (as you will see, this is challenging!). Let $\eta_1(x_0) = \beta_{1,0} + \beta_{1,1}x_0$ and $\eta_2(x_0) = \beta_{2,0} + \beta_{2,1}x_0$ for any x_0. We know $\eta_1(\gamma) = \eta_2(\gamma)$ for γ. Their MLE (LSE) becomes

$$\hat{\eta}_1(x_0) = \hat{\beta}_{1,0} + \hat{\beta}_{1,1}x_0, \quad \hat{\eta}_2(x_0) = \hat{\beta}_{2,0} + \hat{\beta}_{2,1}x_0.$$

$\hat{\eta}_1(x_0)$ and $\hat{\eta}_2(x_0)$ are independent and follow normal distributions. So, the distribution of $\hat{\eta}_1(x_0) - \hat{\eta}_2(x_0)$ also follows a normal distribution whose mean and variance are

$$\begin{aligned}
\mathrm{E}(\hat{\eta}_1(x_0) - \hat{\eta}_2(x_0)) &= \beta_{1,0} + \beta_{1,1}x_0 - \beta_{2,0} - \beta_{2,1}x_0, \\
\mathrm{Var}(\hat{\eta}_1(x_0) - \hat{\eta}_2(x_0)) &= \mathrm{Var}(\hat{\eta}_1(x_0)) + \mathrm{Var}(\hat{\eta}_2(x_0)) \\
&= \sigma^2(\boldsymbol{a}^T(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{a} + \boldsymbol{a}^T(\boldsymbol{X}_2^T\boldsymbol{X}_2)^{-1}\boldsymbol{a}) \\
&= \sigma^2\left(\frac{1}{n} + \frac{(\bar{x}_1 - x_0)^2}{\sum_{i=1}^n (x_i - \bar{x}_1)^2} + \frac{1}{m} + \frac{(\bar{x}_2 - x_0)^2}{\sum_{i=1}^m (x_i - \bar{x}_2)^2}\right) \\
&= \sigma^2 H(x_0)
\end{aligned}$$

where $\boldsymbol{\lambda} = [1,\ x_0]^T$. That is,

$$\hat{\eta}_1(x_0) - \hat{\eta}_2(x_0) \sim \mathrm{N}(\eta_1(x_0) - \eta_2(x_0), \sigma^2 H(x_0)).$$

By letting $x_0 = \gamma$, we have

$$\hat{\eta}_1(\gamma) - \hat{\eta}_2(\gamma) \sim \mathrm{N}(0, \sigma^2 H(\gamma)).$$

Thus,

$$\frac{(\hat{\eta}_1(\gamma) - \hat{\eta}_2(\gamma))^2}{\hat{\sigma}^2 H(\gamma)} = \frac{(\hat{\beta}_{1,0} - \hat{\beta}_{1,1}\gamma - \hat{\beta}_{2,0} - \hat{\beta}_{2,1}\gamma)^2}{\hat{\sigma}^2 H(\gamma)} \sim F(1, n+m-4)$$

Thus, the 95% confidence interval for γ is

$$\{\gamma \mid \frac{(\hat{\beta}_{1,0} - \hat{\beta}_{1,1}\gamma - \hat{\beta}_{2,0} - \hat{\beta}_{2,1}\gamma)^2}{\hat{\sigma}^2 H(\gamma)} < F_\alpha(1, n+m-4)\}.$$

This interval should exist (at least the estimate $\hat{\gamma}$ will exist) as long as more than one distinct x value is observed and $\hat{\beta}_{2,1} - \hat{\beta}_{1,1} \neq 0$ (which has probability 1).

4. Two contrasts $\boldsymbol{c}_i^T\hat{\boldsymbol{\beta}}$ and $\boldsymbol{c}_j^T\hat{\boldsymbol{\beta}}$ are said to be orthogonal if $\boldsymbol{c}_i^T\boldsymbol{c}_j = 0$.
Since we are assuming normality, they are independent if $\mathrm{cov}(\boldsymbol{c}_i^T\hat{\boldsymbol{\beta}},\ \boldsymbol{c}_j^T\hat{\boldsymbol{\beta}})$=0. We also know that $\mathrm{cov}(\boldsymbol{c}_i^T\hat{\boldsymbol{\beta}},\ \boldsymbol{c}_j^T\hat{\boldsymbol{\beta}})$=$\sigma^2\boldsymbol{c}_i^T(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{c}_j$ and this is invariant to the choice of generalized inverse

$(\boldsymbol{X}^T\boldsymbol{X})^-$ because the contrasts are estimable. For the balanced one-way ANOVA model, $y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}$ for $i = 1, 2, ..., k$, and $j = 1, 2, ..., n$

$$\boldsymbol{X} = \begin{bmatrix} 1 & \mathbf{1}_n & \mathbf{0} & \cdots & 0 \\ 1 & \mathbf{0} & \mathbf{1}_n & \cdots & \mathbf{0} \\ 1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_n \end{bmatrix}.$$

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} kn & n & n & \cdots & n \\ n & n & \mathbf{0} & \cdots & \mathbf{0} \\ n & \mathbf{0} & n & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n & \mathbf{0} & \mathbf{0} & \cdots & n \end{bmatrix}.$$

$$(\boldsymbol{X}^T\boldsymbol{X})^- = diag[0, (1/n), ..., (1/n)]$$

and therefore $\mathrm{cov}(\boldsymbol{c}_i^T\hat{\beta},\ \boldsymbol{c}_j^T\hat{\beta}) = \sigma^2\boldsymbol{c}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{c}_j = 0$ if $\boldsymbol{c}_i^T\boldsymbol{c}_j = 0$ (assuming that the first element of $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ are 0).

5. Use the cell means model $y_{i,j} = \mu_i + \epsilon_{i,j}$ with $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^T$ and

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_k} \end{bmatrix}.$$

Then

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_k \end{bmatrix} \text{ and } (\boldsymbol{X}^T\boldsymbol{X})^{-1} = \mathrm{diag}(\frac{1}{n_1}, \frac{1}{n_2}, \ldots, \frac{1}{n_k}).$$

This yields

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = (\bar{y}_{1,\cdot}, \bar{y}_{2,\cdot}, \ldots, \bar{y}_{k,\cdot})^T, \tag{1}$$

with

$$\hat{\boldsymbol{\mu}} \sim \mathrm{N}_k\left(\boldsymbol{\mu}, \mathrm{diag}(\frac{1}{n_1}, \frac{1}{n_2}, \ldots, \frac{1}{n_k})\right). \tag{2}$$

Now by (1) we have $\hat{\delta} = \sum_{i=1}^k a_i\bar{y}_{i,\cdot} = \boldsymbol{a}^T\hat{\boldsymbol{\mu}}$ and $\hat{\gamma} = \sum_{i=1}^k b_i\bar{y}_{i,\cdot} = \boldsymbol{b}^T\hat{\boldsymbol{\mu}}$. Then by normal distribution theory and (2), we know that $\hat{\delta}$ and $\hat{\gamma}$ are independent if and only if

$$\boldsymbol{a}^T\mathrm{Cov}(\hat{\boldsymbol{\mu}})\boldsymbol{b} = \boldsymbol{a}^T \begin{bmatrix} \frac{1}{n_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{n_k} \end{bmatrix} \boldsymbol{b} = \sum_{i=1}^k a_ib_i/n_i = 0. \quad \square$$

6. (a)

$$\begin{aligned}
\mathrm{E}[\hat{\boldsymbol{\epsilon}}] &= \mathrm{E}[(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y}] \\
&= (\boldsymbol{I}-\boldsymbol{P})\mathrm{E}[\boldsymbol{y}] \\
&= (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{X}\boldsymbol{\beta} \\
&= \boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta} \\
&= \boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{X}\boldsymbol{\beta} && (\boldsymbol{P} \text{ projects } \boldsymbol{X} \text{ to } \boldsymbol{X}) \\
&= \boldsymbol{0}\,.
\end{aligned}$$

(b)

$$\begin{aligned}
\mathrm{Cov}[\hat{\boldsymbol{\epsilon}}] &= \mathrm{Cov}[(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y}] \\
&= (\boldsymbol{I}-\boldsymbol{P})\mathrm{Cov}[\boldsymbol{y}](\boldsymbol{I}-\boldsymbol{P})^T \\
&= (\boldsymbol{I}-\boldsymbol{P})\sigma^2\boldsymbol{I}(\boldsymbol{I}-\boldsymbol{P})^T \\
&= \sigma^2(\boldsymbol{I}-\boldsymbol{P})(\boldsymbol{I}-\boldsymbol{P}) && ((\boldsymbol{I}-\boldsymbol{P}) \text{ is symmetric}) \\
&= \sigma^2(\boldsymbol{I}-\boldsymbol{P})\,. && ((\boldsymbol{I}-\boldsymbol{P}) \text{ is idempotent})
\end{aligned}$$

(c)

$$\begin{aligned}
\mathrm{Cov}[\hat{\boldsymbol{\epsilon}},\boldsymbol{y}] &= \mathrm{Cov}[(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y},\boldsymbol{y}] \\
&= (\boldsymbol{I}-\boldsymbol{P})\mathrm{Cov}[\boldsymbol{y},\boldsymbol{y}] \\
&= (\boldsymbol{I}-\boldsymbol{P})\sigma^2\boldsymbol{I} \\
&= \sigma^2(\boldsymbol{I}-\boldsymbol{P})\,.
\end{aligned}$$

(d)

$$\begin{aligned}
\mathrm{Cov}[\hat{\boldsymbol{\epsilon}},\hat{\boldsymbol{y}}] &= \mathrm{Cov}[(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y},\boldsymbol{P}\boldsymbol{y}] \\
&= (\boldsymbol{I}-\boldsymbol{P})\mathrm{Cov}[\boldsymbol{y},\boldsymbol{y}]\boldsymbol{P}^T \\
&= (\boldsymbol{I}-\boldsymbol{P})\sigma^2\boldsymbol{I}\boldsymbol{P}^T \\
&= \sigma^2(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{P} && (\boldsymbol{P} \text{ is symmetric}) \\
&= \sigma^2(\boldsymbol{P}-\boldsymbol{P}\boldsymbol{P}) \\
&= \sigma^2(\boldsymbol{P}-\boldsymbol{P}) && (\boldsymbol{P} \text{ is idempotent}) \\
&= \boldsymbol{0}\,.
\end{aligned}$$